A List of countries and sectors

Tab.10 contains the list of the main EU-15 trade partners in 1995-2009. In blue are the six countries dropped from the sample because they are absent from the WIOD database.

Table 10: List of main EU-15 trading partners

ID	Country	Type	ID	Country	Type
AT	Austria	intra-eu15	PL	Poland	ceec
BE	Belgium-Luxembourg	intra-eu15	RO	Romania	ceec
DK	Denmark	intra-eu15	SK	Slovakia	ceec
FI	Finland	intra-eu15	SI	Slovenia	ceec
FR	France	intra-eu15	TR	Turkey	ceec
DE	Germany	intra-eu15	CA	Canada	other devpd
GR	Greece	intra-eu15	JP	Japan	other devpd
IE	Ireland	intra-eu15	KR	Korea	other devpd
IT	Italy	intra-eu15	NO	Norway	other devpd
NL	Netherlands	intra-eu15	CH	Switzerland	other devpd
PT	Portugal	intra-eu15	US	USA	other devpd
ES	Spain	intra-eu15	BR	Brazil	other emerging
SW	Sweden	intra-eu15	CN	China	other emerging
GB	United Kingdom	intra-eu15	IN	India	other emerging
BG	Bulgaria	ceec	ID	Indonesia	other emerging
HR	Croatia	ceec	MY	Malaysia	other emerging
CZ	Czech Republic	ceec	MX	Mexico	other emerging
EE	Estonia	ceec	RU	Russia	other emerging
HU	Hungary	ceec	SG	Singapore	other emerging
LV	Latvia	ceec	TW	Taiwan	other emerging
LT	Lithuania	ceec	TH	Thailand	other emerging

The indicator of proximity endowment can be constructed for all countries because data on bilateral distance is widely available. Tab.11 provides details on the sample of countries for which we can also construct domestic supply based proximity indicators (DSBP) derived in sec.. In black are the countries for which output data is consistent with trade data, and for which we can construct the DSBP indicator. In blue are the countries for which output data is either lacking or inconsistent with trade data. For such countries we can only construct the indicator of foreign supply based proximity which only takes into account the distribution of market shares across foreign suppliers. In red is the only country for which no bilateral imports data is available (Taiwan). In the empirical analysis we impute Taiwan's proximity characteristic using the indicator of proximity endowment and the estimated relationship of DSBP with proximity endowment. We do not construct DSBP indicators for countries in blue using this relationship because these countries drop out anyway in the second step of the estimation.[55]

[55] Either measured TFP data is of poor quality or sectoral R&D data is missing, or both.

Table 11: Proximity indicators in the sample

ID	Country	Type	ID	Country	Type
AT	Austria	intra-eu15	PL	Poland	ceec
BE	*Belgium-Luxembourg*	intra-eu15	RO	Romania	ceec
DK	Denmark	intra-eu15	SK	Slovakia	ceec
FI	Finland	intra-eu15	SI	Slovenia	ceec
FR	France	intra-eu15	TR	Turkey	ceec
DE	Germany	intra-eu15	CA	Canada	other devpd
GR	Greece	intra-eu15	JP	Japan	other devpd
IE	Ireland	intra-eu15	KR	Korea	other devpd
IT	Italy	intra-eu15	NO	*Norway*	other devpd
NL	*Netherlands*	intra-eu15	CH	*Switzerland*	other devpd
PT	Portugal	intra-eu15	US	USA	other devpd
ES	Spain	intra-eu15	BR	Brazil	other emerging
SW	Sweden	intra-eu15	CN	China	other emerging
GB	United Kingdom	intra-eu15	IN	India	other emerging
BG	Bulgaria	ceec	ID	*Indonesia*	other emerging
HR	*Croatia*	ceec	MY	*Malaysia*	other emerging
CZ	Czech Republic	ceec	MX	Mexico	other emerging
EE	Estonia	ceec	RU	Russia	other emerging
HU	Hungary	ceec	SG	*Singapore*	other emerging
LV	Latvia	ceec	TW	*Taiwan*	other emerging
LT	Lithuania	ceec	TH	*Thailand*	other emerging

Table 12: Input Intensity in Production and Trade (EU-15)

ID	Desc	ii_{prod}'95	ii_{prod}'09	ii_{trade}'95	ii_{trade}'09
15 – 16	Manuf. Food-Tobacco	.75	.76	.22	.22
17 – 18	Manuf. Textile-Clothes	.66	.69	.25	.14
19	Manuf. Leather	.71	.70	.14	.05
20	Manuf. Wood (no Furniture)	.66	.70	.79	.77
21 – 22	Manuf. Paper and Publishing (Media)	.62	.65	.74	.76
24	Manuf. Chemicals-Pharmaceuticals	.66	.71	.84	.67
25	Manuf. Rubber-Plastic	.63	.68	.70	.67
26	Manuf. Non-Metallic Products	.59	.65	.78	.81
27 – 28	Manuf. Metal Products (no Machinery)	.65	.70	.90	.87
29	Manuf. Machinery-Equipment	.63	.66	.38	.44
30 – 33	Manuf. 'Other' Equipment	.64	.68	.48	.33
34 – 35	Manuf. Transport Equipment	.73	.79	.42	.41
36 – 37	Misc. Manuf. (furniture, toys,...)	.64	.67	.24	.17

ii_{trade} (ii_{prod}) is input intensity in trade (resp., production). Input intensity in production is the income share of inputs in gross output constructed for the EU-15 aggregate. Input intensity in trade is the value share of processed intermediate inputs in total imports constructed for EU-15 trade with non-EU15 (excl. intra-EU).
Primary and processed inputs in the oil industry are excluded from calculation of ii_{trade}. 'Other' is manufacturing of computers, electrical, audiovisual, communications, measurement, and precision equipment.

B Use of inputs and the level of aggregation

Tab.13 provides information on the share of inputs in production (ii_{prod}) for 22 manufacturing sectors which correspond to the 2-digit level of the ISIC Rev.3 classification. We focus on the main economies of the EU-15 in 1996-2006, i.e. Germany, France, UK, Italy, Spain, Finland, and Sweden, because the UNIDO INDSTAT database provides consistent time series data on gross output ($PROD$) and value added (VA) at the 4-digit level in the ISIC rev.3

for these countries.[56]

The data is judged consistent in a given year if output data is consistent with data on total exports and imports. This is the case when consumption of varieties produced domestically is a positive fraction of total sectoral expenditure, and this fraction is relatively persistent in adjacent years. The income share of inputs in sectoral production is computed as $ii_{prod} = 1 - VA/PROD$. Country specific input intensity indicators are aggregated into an EU7 indicator according to countries' share in total sectoral output.[57]

Table 13: Inputs' share in gross output (EU7), 2-digit level (ISIC Rev.3)

ID	Description	Ranking	Share 1996	Share 2006
15	Manufacture of food products and beverages	3	.77	.78
16	Manufacture of tobacco products	3	.89	.88
17	Manufacture of textiles	2	.69	.72
18	Manufacture of wearing apparel	2	.66	.73
19	Tanning and dressing of leather	2	.71	.75
20	Manufacture of wood products, except furniture	2	.69	.71
21	Manufacture of pulp, paper, paper products	2	.65	.71
22	Publishing, printing, reproduction of media	1	.62	.65
23	Manufacture of coke, refined petroleum products, nuclear fuel	3	.90	.92
24	Manufacture of chemicals and chemical products	2	.68	.71
25	Manufacture of rubber & plastic products	1	.63	.70
26	Manufacture of other non-metallic mineral products	1	.61	.67
27	Manufacture of basic metals	3	.74	.79
28	Manufacture of fabricated metal products, except machinery	1	.62	.66
29	Manufacture of machinery & equipment	2	.65	.68
30	Manufacture of office machinery & computers	3	.75	.75
31	Manufacture of electrical machinery & apparatus	1	.64	.69
32	Manufacture of radio, TV, communication equipment	2	.69	.71
33	Manufacture of medical, precision & optical instruments	1	.59	.59
34	Manufacture of motor vehicles & trailers	3	.77	.79
35	Manufacture of other transport equipment	2	.72	.71
36	Manufacture of furniture	2	.66	.69

In column 3, '1' corresponds to ii_{prod} below iqr, '2' corresponds to ii_{prod} in the iqr, and '3' corresponds to ii_{prod} above iqr for the EU7 aggregate.

The objective of this section is to check whether the assumption of the model that input intensity is a sector specific characteristic common across countries and invariant overtime is tenable. Further, we want to verify how sensitive are the indicators of sectoral input intensity to the level of aggregation of the data. Tab.13 illustrates that input intensity in production is a persistent characteristic of the sector.[58] Persistence is also verified at the

[56] There is no data on Germany in 1996-1997, and no data on UK in 2006.

[57] There are two bottlenecks in extending the dataset to more years or countries: availability of VA data, consistency of domestic absorption measures overtime.

[58] At the 4-digit level for the EU7 as an aggregate, the correlation coefficient between input intensity measures in 1996 and 2001 is .91, and .88 with 2006. At the 2-digit level, the corresponding measures are .97 and .96.

level of individual countries, and reinforced for individual countries when we work at the 4-digit level.

This table also documents substantial intersectoral variability in the income share of inputs in production. In particular, if the indicators are computed for the EU7 as an aggregate, within-sectoral input intensity variability is lower than intersectoral variability. Within-sectoral interquartile range varies between .01 and .05 for the 21 manufacturing sectors while the standard deviation of sector specific median ii_{prod} indicators is .06.

Table 14 and fig.12 show that the share of inputs in production is relatively homogeneous across these 7 countries. In particular, it is possible to establish an ordinal ranking of sectors according to their input intensity which is relatively stable across countries and overtime. This pattern is particularly strong for the 5 biggest EU15 economies which are Germany, France, Italy, Spain, and the UK for which the correlation coefficient of sectoral ii_{prod} varies between .84-.93.

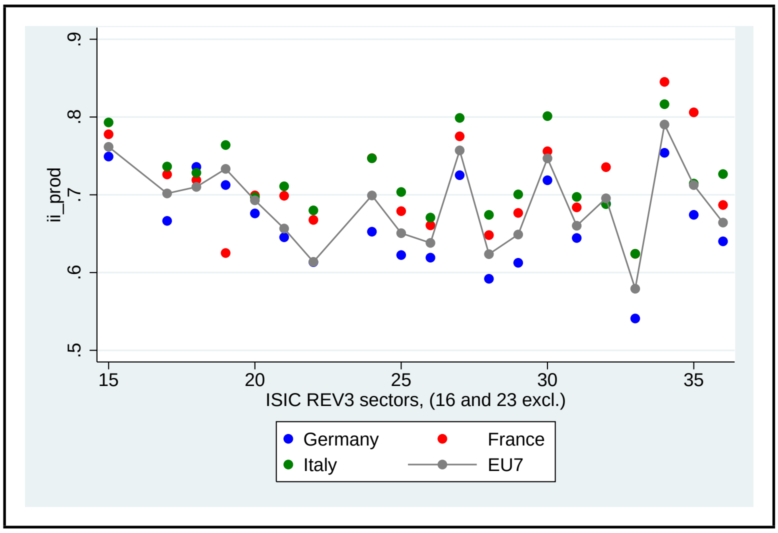
Table 14: Input intensity in production in EU7 in 1996-2006

sector	EU7	GERMANY	FRANCE	UK	ITALY	SPAIN	FINLAND	SWEDEN
15	0.76	0.75	0.78	0.69	0.79	0.79	0.74	0.75
17	0.70	0.67	0.73	0.61	0.74	0.70	0.58	0.63
18	0.71	0.74	0.72	0.58	0.73	0.68	0.62	0.69
19	0.73	0.71	0.63	0.58	0.76	0.75	0.61	0.67
20	0.69	0.68	0.70	0.61	0.70	0.70	0.76	0.74
21	0.66	0.65	0.70	0.59	0.71	0.66	0.68	0.66
22	0.61	0.61	0.67	0.53	0.68	0.61	0.59	0.67
23	0.92	0.94	0.93	0.92	0.93	0.83	0.86	0.71
24	0.70	0.65	0.75	0.62	0.75	0.71	0.67	0.54
25	0.65	0.62	0.68	0.60	0.70	0.66	0.59	0.63
26	0.64	0.62	0.66	0.57	0.67	0.64	0.61	0.64
27	0.76	0.73	0.78	0.75	0.80	0.76	0.75	0.72
28	0.62	0.59	0.65	0.57	0.67	0.65	0.62	0.61
29	0.65	0.61	0.68	0.61	0.70	0.65	0.69	0.65
30	0.75	0.72	0.76	0.76	0.80	0.78	0.80	0.67
31	0.66	0.64	0.68	0.61	0.70	0.70	0.67	0.68
32	0.70	0.69	0.74	0.66	0.69	0.74	0.65	0.77
33	0.58	0.54	0.62	0.54	0.62	0.61	0.60	0.59
34	0.79	0.75	0.85	0.77	0.82	0.81	0.63	0.77
35	0.71	0.67	0.81	0.59	0.71	0.69	0.71	0.64
36	0.66	0.64	0.69	0.59	0.73	0.66	0.61	0.66

The table gives the median value of ii_{prod} by sector and country in 1996-2006.

There are two main drawbacks in working with these ii_{prod} indicators. First, intersectoral variability in the share of processed inputs in production is blurred by inclusion of raw inputs. This caveat is present at any level of disaggregation. Second, the stability of the input intensity measure across countries is blurred by the variability of the product mix which countries

Figure 12: Input intensity in production: ordinal rankings across EU7



produce within any given 2-digit sector. For the EU7 taken as an aggregate, the problem disappears because the measure is dominated by countries such as Germany, France, and UK for which within-sectoral variability is relatively low and persistence overtime very high.

But at the individual country level, variability in within-sector product mixes quickly becomes a problem, both in terms of sectoral homogeneity and stability overtime. Indeed, even if the hypothesis of ii_{prod} being a sector-specific characteristic invariant across countries may hold in terms of ordinal sector rankings, it fits the data better at the level of 121 manufacturing sectors at the 4-digit level.

To illustrate we look at a simple indicator. At the 2-digit level, the correlation coefficient between country-specific input intensity measures in 1998 varies between .82 for Germany and Italy and .34 for Germany and Sweden. In 2005 the correlation coefficient for Germany-Italy is virtually unchanged while for Germany and Sweden it jumps up to .83.[59]

At the 4-digit level this problem is alleviated. The correlation coefficient between country-specific vectors of ii_{prod} varies between .76 (.77) for Germany-France and .64 (.69) for Germany-Sweden in 1998 (2005). Cross-country and overtime variability are strongly reduced. Furthermore, ii_{prod} indicators for Germany are more strongly correlated to France than to Italy,

[59]This example voluntarily looks at extremes: variability is much lower for the five main EU-15 economies as illustrated by the Italy-Germany measure.

contrary to the input intensity measure constructed at the 2-digit level.

We conclude that available evidence points to differences in within-sectoral specialization across EU-15 countries rather than to intrinsic differences in production functions. This may lead to difficulties in using ii_{prod} measures constructed at the 2-digit level in empirical analysis. Nonetheless, available empirical evidence at the 4-digit level is consistent with the hypothesis of input intensity in production being a sector-specific characteristic invariant across countries. This assumption is built into the model.

C Revealed Comparative Advantage

This appendix provides descriptive statistics on estimated exporter-sector dummies for 20 manufacturing sectors which correspond to the 2-digit level in the NACE 1.1 (ISIC Rev.3) classification. In Costinot et al. (2012), this would correspond to fundamental sectoral productivity estimates while in this paper the sectoral cost comprises technology, factor cost, input cost, and export-side trade cost components specific to this sector and exporter.

The fixed effects approach in the sectoral exports regression is implemented at the bilateral level as well as by partner types. The four partner types are: 'intra-eu15' which includes the 15 EU members by 1995, 'ceec' which includes 12 European emerging economies,[60] 'other emerging' which includes the main non-EU emerging economies such as Brazil, China, and India, and 'other developed' which includes the main non-EU members of the OECD such as the United States and Japan.[61]

Tables 15 and 16 give estimated values of relative sectoral exporter-sector dummies in 1995 and 2010 by trade partner group. In Costinot et al. (2012) this would correspond to fundamental sectoral productivity relatively to the processed foods' sector (15) and the group of 'other developed countries' where the exponential of the estimated dummy is normalized to 1. In the model used in this paper, the dummy contains technology, wage, proximity, and export-side trade cost components. The variance in sectoral costs has increased overtime for the group of non-European emerging economies while it has remained stable for the EU-15 and the CEECs.

Tables 17 and 18 provide estimates of relative exporter specific sectoral production costs for a subset of countries in the bilateral estimation. Cost

[60] This group is constituted by the 10 Central and Eastern European countries which have become members of the European Union in 2004 and 2007, Turkey which has a customs union with the EU since 1996, and Croatia which is set to become the 28th member of the EU on Jan.1, 2013. Cyprus and Malta are not included in the analysis.

[61] The grouping of 42 main EU-15 partners in 4 types is detailed in table (10).

Table 15: Exporter-sector production costs by partner type in 1995

Sector	**ceec**	**non-EU emerging**	**intra-eu15**	**other devpd**
15	1	1	1	1
17	1.148	1.035	0.972	1
18	1.397	1.307	1.086	1
19	1.221	1.344	1.047	1
20	1.135	1.061	0.885	1
21	0.852	0.772	0.926	1
22	0.73	0.73	0.87	1
24	0.872	0.784	0.879	1
25	0.929	0.869	0.919	1
26	1.174	0.904	1	1
27	1.089	0.934	0.931	1
28	1	0.917	0.939	1
29	0.847	0.736	0.849	1
30	0.576	0.91	0.816	1
31	0.895	0.856	0.84	1
32	0.762	0.892	0.791	1
33	0.626	0.723	0.741	1
34	0.876	0.663	0.932	1
35	0.664	0.697	0.693	1
36	0.975	1.063	0.873	1

The reported relative cost corresponds to $\left[e^{fe^k_{i,t}}\right]^{1/\theta}$.

Table 16: Exporter-sector production costs by partner type in 2010

Sector	**ceec**	**non-EU emerging**	**intra-eu15**	**other devpd**
15	1	1	1	1
17	1.217	1.274	0.985	1
18	1.471	1.619	1.201	1
19	1.211	1.606	1.194	1
20	1.149	1.102	0.917	1
21	0.919	0.878	0.921	1
22	0.923	0.963	0.949	1
24	0.776	0.813	0.84	1
25	1.036	0.998	0.924	1
26	1.102	1.097	0.968	1
27	0.915	0.922	0.87	1
28	1.061	1.066	0.924	1
29	0.934	0.912	0.843	1
30	1.093	1.154	0.939	1
31	1.036	1.034	0.877	1
32	1.034	1.1	0.839	1
33	0.73	0.814	0.748	1
34	1.109	0.819	0.962	1
35	0.651	0.792	0.671	1
36	1.113	1.246	0.927	1

The reported relative cost corresponds to $\left[e^{fe^k_{i,t}}\right]^{1/\theta}$.

is normalized to 1 in the US in all sectors, and in the food sector for all countries. There is substantial heterogeneity in RCA rankings within each partner group.

Table 17: Exporter-sector production costs for a subsample of countries in 1995

sector	CZECH	HUNGARY	CHINA	MEXICO	CANADA	JAPAN	GERMANY	ITALY
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
15	1	1	1	1	1	1	1	1
17	1.446	1.07	1.42	0.969	0.925	1.59	1.131	1.255
18	1.378	1.174	1.846	0.784	0.936	1.188	1.138	1.282
19	1.465	1.083	1.856	0.956	0.811	1.204	1.026	1.43
20	1.2	0.806	1.129	0.534	1.182	0.779	0.884	0.813
21	1.142	0.732	0.809	0.561	1.298	1.218	1.054	0.919
22	0.909	0.721	0.888	0.545	0.928	1.226	0.881	0.788
24	1.042	0.885	0.986	0.895	0.808	1.584	1.007	0.878
25	1.174	0.972	1.191	0.734	0.837	1.751	1.072	1.044
26	1.568	1.045	1.249	0.908	0.97	1.674	1.122	1.177
27	1.298	1.11	0.957	0.996	1.192	1.424	1.15	1.078
28	1.289	0.995	1.311	0.745	0.912	1.651	1.137	1.113
29	1.141	0.848	0.99	0.653	0.872	1.771	1.014	1.018
30	0.668	0.468	0.999	0.717	0.811	1.602	0.827	0.727
31	1.063	0.975	1.175	0.815	0.822	1.811	1.02	0.907
32	0.753	0.877	1.113	0.78	0.856	1.886	0.92	0.766
33	0.739	0.601	0.957	0.657	0.788	1.595	0.848	0.731
34	1.302	1.031	0.629	0.647	0.89	2.32	1.296	1.094
35	0.771	0.408	0.669	0.548	0.885	1.386	0.676	0.671
36	1.261	0.862	1.542	0.954	0.912	1.73	0.972	1.096

Table 19 provides information on the persistence of sectoral rankings overtime at the level of 2-digit sectors for each group of partners. The vector of correlation coefficients for cost rankings in 1995 with subsequent years is reported for each partner type. Cost rankings are strongly persistent in all country groups with the exception of the CEECs. On the other hand, non-European emerging economies are characterized by very persistent cost rankings.

Table 20 provides information on the persistence of sectoral cost rankings overtime at the level of 2-digit sectors for a subsample of countries. The table reports the vector of correlation coefficients for cost rankings in 1995 with subsequent years for 2 countries by partner type.

There is strong variability in relative sectoral production costs for the CEECs in 1995-2010. For the group of non European emerging economies results for China and Mexico are reported. China is a typical country in the group of non-European emerging economies in the sense that the ranking of its sectoral costs if very persistent overtime. Mexico is the only non-EU emerging economy which shows substantial changes in revealed comparative

Table 18: Exporter-sector production costs for a subsample of countries in 2010

sector	CZECH	HUNGARY	CHINA	MEXICO	CANADA	JAPAN	GERMANY	ITALY
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
15	1	1	1	1	1	1	1	1
17	1.218	1.013	1.621	0.766	0.858	1.474	1.022	1.148
18	1.289	1.024	1.972	0.927	0.986	1.116	1.121	1.265
19	1.204	0.925	1.937	0.982	0.785	1.084	1.044	1.361
20	0.909	0.927	1.202	0.471	1.055	0.741	0.89	0.758
21	1.005	0.834	0.957	0.631	0.958	1.128	0.918	0.849
22	1.023	0.763	1.099	0.632	0.887	1.186	0.889	0.812
24	0.826	0.762	0.867	0.783	0.798	1.34	0.839	0.777
25	1.145	1.032	1.204	0.746	0.828	1.574	0.969	0.947
26	1.114	0.965	1.29	0.845	0.793	1.394	0.937	0.974
27	1.05	0.869	1.06	0.765	0.931	1.431	0.985	0.958
28	1.18	0.962	1.312	0.708	0.874	1.469	0.989	0.967
29	0.999	0.936	1.085	0.778	0.846	1.559	0.868	0.891
30	1.354	1.084	1.333	0.922	0.87	1.451	0.89	0.666
31	1.129	1.067	1.261	0.943	0.846	1.542	0.937	0.862
32	1.138	1.23	1.347	1.043	0.879	1.562	0.86	0.682
33	0.747	0.757	0.895	0.863	0.784	1.361	0.733	0.646
34	1.553	1.267	1.011	0.945	0.851	2.14	1.242	1.033
35	0.643	0.527	0.829	0.5	0.902	1.18	0.647	0.637
36	1.16	0.888	1.509	0.841	0.913	1.434	0.929	0.956

Table 19: Rankings' autocorrelation by partner type

Year	eu15	ceec	emerging
1998	0.98	0.95	0.99
1999	0.97	0.96	0.99
2000	0.96	0.94	0.99
2001	0.96	0.93	0.99
2002	0.95	0.93	0.97
2003	0.95	0.91	0.98
2004	0.95	0.89	0.98
2005	0.94	0.85	0.99
2006	0.94	0.83	0.98
2007	0.92	0.8	0.98
2008	0.92	0.75	0.97
2009	0.9	0.72	0.95
2010	0.9	0.71	0.95

advantage on EU-15 markets at the level of 2-digit sectors. The period of study corresponds to the implementation of NAFTA, i.e. Mexico undergoes a process of regional integration with the Canadian and US economies which may impact the ranking of sectoral production costs for Mexico on world markets.

Table 20 indicates strong variability in revealed relative sectoral produc-

Table 20: Rankings' correlation for subsample of countries

sector	CANADA	CHINA	CZECH	GERMANY	HUNGARY	ITALY	JAPAN	MEXICO
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
1998	0.97	0.99	0.94	0.98	0.84	0.99	0.98	0.91
1999	0.96	0.99	0.93	0.97	0.82	0.99	0.98	0.88
2000	0.96	0.98	0.93	0.95	0.77	0.99	0.98	0.70
2001	0.96	0.98	0.91	0.94	0.78	0.98	0.98	0.68
2002	0.93	0.98	0.89	0.92	0.80	0.98	0.96	0.59
2003	0.94	0.97	0.86	0.92	0.75	0.98	0.98	0.56
2004	0.92	0.96	0.80	0.92	0.67	0.99	0.97	0.60
2005	0.91	0.96	0.71	0.91	0.60	0.99	0.98	0.67
2006	0.89	0.96	0.68	0.92	0.57	0.99	0.97	0.61
2007	0.94	0.95	0.64	0.89	0.53	0.98	0.98	0.64
2008	0.92	0.95	0.60	0.89	0.51	0.98	0.98	0.56
2009	0.87	0.94	0.56	0.89	0.54	0.98	0.97	0.55
2010	0.84	0.92	0.51	0.88	0.45	0.97	0.97	0.62

tion costs for the CEECs in 1995-2010. One explanation could be changes in data reporting thresholds for intra-EU trade which have impacted 3/4 of the CEEC group in 2004 and 1/6 of the group in 2007 upon entry in the European Union. As shown in table 21, a simple regression of cost correlation coefficients at the country level confirms there is a break in 2004 for the CEECs. However, as shown in col.3, the most robust feature of the data is the stronger variability in rankings for the CEECs relatively to other partner types over the whole period.

Tables 22 and 23 report annual ratios of revealed relative sectoral costs for the CEECs relatively to respectively non-European emerging economies and the countries of the EU-15. Sectors are ranked according to RCA in 1995. The theoretical underpinning of these tables is the prediction of the Ricardian model that the ratio of observed relative bilateral exports is increasing in the gap of fundamental sectoral costs, controlling for pair and destination-sector fixed effects.

These tables show that the pattern of comparative advantage for the CEECs on EU-15 markets varies across partner groups. To give an example, the CEECs have a persistent comparative advantage in the wearing apparel industry (18) relatively to the EU-15, but a consistent disadvantage relatively to non-European emerging economies. Another example is the motor vehicles industry (34) in which the CEECs have a persistent strong comparative advantage relatively to non-European emerging economies while this industry is not in the top 5 for the CEECs relatively to the EU-15.

Several features of the data stand out. First, there is evidence of substantial modification in cost rankings as well as in the RCA pattern for the

Table 21: **Persistence of cost rankings overtime**

depvar: Correlation coefficient	(1)	(2)	(3)
year	-0.0138***	-0.0076***	-0.0085***
	(0.0032)	(0.0017)	(0.0033)
*year*intraeu15*		0.0006	0.0003
		(0.0019)	(0.0038)
*entry*ceec*	-0.0103*		-0.000
	(0.0058)		(0.000)
*year*ceec*		-0.0104***	-0.0085**
		(0.0020)	(0.0039)
*year*emerg*		-0.0003	-0.0013
		(0.0020)	(0.0040)
*entry*intraeu15*			0.000
			(0.000)
*entry*emerg*			0.000
			(0.000)
Entry FE	YES	NO	YES
Type FE		YES	YES
Observations	192	656	656
R-squared	0.484	0.446	0.448

The dependent variable is the correlation coefficient of the sectoral production cost in 1995 with every other year.
'entry' is a dummy equal to 1 in 2004-2010. *** p<0.01, ** p<0.05, * p<0.1.
'year' picks up the time trend. In (1) the regression is on the CEEC sample.
In (2)-(3), the regression is on the full sample of EU-15 trade partners.

CEECs on EU-15 markets relatively to all partner groups, and in particular relatively to non-EU emerging economies. Second, the variance of cost ratios at the intersectoral level is reduced overtime for CEECs relatively to all partner groups while the gap in extreme values is not. The max-min gap increases for CEECs relatively to non-EU emerging economies while it is reduced relatively to the EU-15 country group.

Table 24 reports the pattern of revealed comparative advantage for the EU-15 relatively to non-EU emerging economies.

D Addenda on Estimation of the Model

D.1 R&D data: imputation of missing observations

Data on sectoral R&D activity is taken from the 2011 edition of OECD ANBERD for all countries but China.[62] For most countries, a fair amount of

[62] Seven countries are dropped from the analysis for lack of R&D data. These are Bulgaria, Brazil, India, Indonesia, Lithuania, Latvia, and Russia.

Table 22: Specialization pattern of CEECs to non-EU emerging economies

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
34	1.32	1.36	1.36	1.39	1.36	1.46	1.43	1.47	1.48	1.50	1.45	1.44	1.42	1.40	1.36	1.36
26	1.30	1.30	1.32	1.29	1.24	1.24	1.24	1.24	1.21	1.16	1.11	1.08	1.04	1.04	1.02	1.00
23	1.28	1.50	1.32	1.00	1.17	1.09	0.99	0.94	0.94	0.91	0.99	0.98	1.09	1.09	1.03	1.14
27	1.17	1.16	1.19	1.19	1.19	1.17	1.18	1.16	1.12	1.10	1.05	1.04	1.02	1.02	1.03	0.99
29	1.15	1.15	1.17	1.18	1.14	1.15	1.14	1.15	1.13	1.10	1.08	1.08	1.08	1.08	1.05	1.02
24	1.11	1.12	1.12	1.10	1.06	1.10	1.06	1.05	1.02	1.00	0.97	0.97	0.96	0.97	0.97	0.95
17	1.11	1.11	1.13	1.15	1.13	1.15	1.15	1.14	1.14	1.12	1.07	1.07	1.05	1.02	0.98	0.96
21	1.10	1.16	1.21	1.18	1.15	1.18	1.16	1.15	1.14	1.14	1.11	1.12	1.11	1.09	1.07	1.05
28	1.09	1.08	1.10	1.11	1.08	1.12	1.11	1.11	1.12	1.08	1.06	1.05	1.05	1.04	1.03	1.00
25	1.07	1.07	1.08	1.09	1.07	1.10	1.11	1.12	1.13	1.10	1.08	1.07	1.07	1.06	1.05	1.04
20	1.07	1.10	1.14	1.14	1.12	1.15	1.13	1.14	1.13	1.10	1.07	1.07	1.06	1.04	1.04	1.04
18	1.07	1.09	1.11	1.13	1.11	1.13	1.13	1.10	1.09	1.07	1.01	0.98	0.98	0.95	0.92	0.91
31	1.05	1.07	1.09	1.09	1.06	1.08	1.10	1.10	1.09	1.07	1.05	1.06	1.06	1.04	1.02	1.00
22	1.00	1.03	1.05	1.04	1.00	1.04	1.06	1.05	1.06	1.01	1.01	1.05	1.02	0.98	0.98	0.96
35	0.95	0.98	0.95	0.93	0.92	0.96	0.97	1.06	0.96	1.00	0.90	0.93	0.91	0.90	0.88	0.82
36	0.92	0.92	0.95	0.96	0.94	0.96	0.97	0.98	0.98	0.97	0.94	0.94	0.93	0.93	0.91	0.89
19	0.91	0.88	0.90	0.90	0.88	0.90	0.90	0.91	0.91	0.88	0.83	0.81	0.80	0.78	0.76	0.75
33	0.87	0.87	0.90	0.91	0.91	0.94	0.93	0.94	0.95	0.94	0.91	0.91	0.92	0.92	0.91	0.90
32	0.86	0.90	0.98	1.03	1.01	1.04	1.05	1.03	1.01	1.03	0.99	0.98	1.00	0.99	0.98	0.94
30	0.63	0.66	0.74	0.80	0.78	0.83	0.81	0.80	0.82	0.82	0.90	0.91	0.95	0.96	0.97	0.95

Table 23: Specialization pattern of CEECs to intra-EU15

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
18	1.29	1.31	1.35	1.38	1.39	1.42	1.42	1.42	1.41	1.38	1.32	1.30	1.28	1.26	1.23	1.23
20	1.28	1.30	1.35	1.38	1.39	1.42	1.40	1.42	1.43	1.36	1.32	1.31	1.28	1.25	1.24	1.25
17	1.18	1.19	1.23	1.27	1.27	1.30	1.31	1.33	1.32	1.31	1.28	1.28	1.27	1.26	1.24	1.24
26	1.18	1.18	1.21	1.23	1.23	1.26	1.25	1.27	1.26	1.23	1.19	1.18	1.16	1.13	1.13	1.14
27	1.17	1.13	1.16	1.20	1.18	1.21	1.20	1.20	1.19	1.18	1.13	1.14	1.12	1.10	1.07	1.05
19	1.17	1.13	1.15	1.15	1.14	1.18	1.18	1.19	1.20	1.17	1.12	1.09	1.06	1.04	1.01	1.01
36	1.12	1.12	1.15	1.17	1.17	1.21	1.22	1.27	1.29	1.27	1.24	1.25	1.23	1.22	1.19	1.20
31	1.07	1.10	1.13	1.16	1.16	1.21	1.22	1.25	1.25	1.23	1.21	1.22	1.22	1.20	1.18	1.18
28	1.06	1.06	1.10	1.13	1.12	1.18	1.18	1.20	1.23	1.18	1.17	1.17	1.17	1.17	1.16	1.15
25	1.01	1.01	1.03	1.06	1.07	1.11	1.12	1.15	1.16	1.13	1.11	1.12	1.12	1.11	1.11	1.12
29	1.00	1.00	1.03	1.05	1.05	1.08	1.09	1.13	1.14	1.12	1.11	1.11	1.11	1.11	1.10	1.11
24	0.99	0.98	0.98	0.98	0.96	1.00	0.97	0.97	0.96	0.93	0.91	0.92	0.91	0.93	0.91	0.92
32	0.96	1.00	1.08	1.14	1.12	1.18	1.19	1.22	1.25	1.27	1.23	1.23	1.24	1.26	1.27	1.23
35	0.96	0.99	0.97	1.01	0.99	1.06	1.04	1.13	1.10	1.13	1.06	1.06	1.02	1.02	1.03	0.97
34	0.94	0.96	1.00	1.05	1.05	1.11	1.12	1.14	1.16	1.15	1.14	1.14	1.14	1.14	1.15	1.15
21	0.92	0.93	0.95	0.97	0.97	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33	0.85	0.85	0.89	0.92	0.93	0.96	0.96	0.98	0.99	0.98	0.95	0.97	0.97	0.98	0.97	0.98
22	0.84	0.87	0.88	0.89	0.88	0.90	0.92	0.92	0.94	0.91	0.91	0.96	0.98	0.96	0.98	0.97
30	0.71	0.74	0.81	0.89	0.87	0.92	0.93	0.93	0.98	0.98	1.05	1.08	1.10	1.14	1.16	1.16

observations is missing at the sectoral level in 1995-2009. For such observations the data is imputed.

When data is missing at the 2-digit level but R&D activity is reported at a higher level of aggregation (1-digit or total manufacturing), we impute sectoral data using the share of the sector in total R&D activity in adjacent years. For three countries R&D activity is never reported for certain industries. This is the case of sectors $17-22$ for Great Britain, sectors $17-19$ for Japan, and $19-22$ for Slovakia. For these countries, sectoral data is imputed using data on the share of the sector in R&D activity in the most similar country in terms of the distribution of R&D activity across sectors. For Great Britain sectoral R&D shares are taken from France. For the Slovak

Table 24: Specialization pattern of EU15 to non-EU emerging economies

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
34	1.41	1.41	1.36	1.33	1.30	1.32	1.29	1.28	1.28	1.30	1.27	1.26	1.25	1.23	1.19	1.18
21	1.20	1.25	1.27	1.22	1.18	1.18	1.17	1.15	1.14	1.15	1.11	1.13	1.11	1.10	1.07	1.05
22	1.19	1.18	1.19	1.18	1.15	1.15	1.15	1.15	1.13	1.11	1.11	1.10	1.04	1.03	1.00	0.99
29	1.15	1.14	1.13	1.12	1.08	1.06	1.04	1.02	1.00	0.98	0.98	0.97	0.97	0.97	0.95	0.92
24	1.12	1.15	1.14	1.12	1.11	1.11	1.09	1.09	1.07	1.07	1.07	1.06	1.06	1.05	1.07	1.03
26	1.11	1.10	1.09	1.05	1.01	0.98	0.99	0.97	0.96	0.95	0.93	0.92	0.90	0.91	0.91	0.88
25	1.06	1.06	1.05	1.02	1.00	0.99	0.99	0.98	0.97	0.98	0.97	0.96	0.96	0.95	0.95	0.93
28	1.02	1.02	1.00	0.98	0.96	0.95	0.94	0.93	0.91	0.91	0.90	0.90	0.90	0.89	0.89	0.87
33	1.02	1.02	1.01	0.99	0.98	0.98	0.97	0.95	0.95	0.96	0.95	0.94	0.95	0.94	0.94	0.92
27	1.00	1.03	1.03	0.99	1.01	0.97	0.98	0.97	0.94	0.93	0.93	0.92	0.91	0.93	0.96	0.94
35	0.99	0.98	0.97	0.92	0.93	0.91	0.93	0.94	0.87	0.89	0.85	0.88	0.89	0.88	0.86	0.85
31	0.98	0.98	0.96	0.94	0.91	0.90	0.91	0.88	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.85
17	0.94	0.94	0.92	0.91	0.89	0.89	0.88	0.86	0.86	0.86	0.84	0.83	0.83	0.81	0.79	0.77
30	0.90	0.89	0.90	0.90	0.89	0.90	0.88	0.86	0.84	0.84	0.85	0.85	0.87	0.84	0.83	0.81
32	0.89	0.90	0.91	0.90	0.90	0.88	0.88	0.85	0.81	0.81	0.81	0.79	0.80	0.79	0.77	0.76
20	0.83	0.84	0.84	0.83	0.81	0.81	0.81	0.80	0.80	0.80	0.81	0.82	0.82	0.83	0.84	0.83
18	0.83	0.83	0.82	0.82	0.80	0.79	0.79	0.78	0.77	0.78	0.76	0.76	0.76	0.76	0.75	0.74
36	0.82	0.83	0.82	0.82	0.80	0.79	0.79	0.77	0.76	0.76	0.76	0.76	0.76	0.77	0.76	0.74
19	0.78	0.78	0.79	0.78	0.77	0.77	0.77	0.76	0.76	0.76	0.74	0.75	0.75	0.76	0.75	0.74

Republic (resp. Japan), sectoral shares are taken from the Czech Republic (resp. South Korea).

When R&D data is missing at higher aggregation levels but reported at the sectoral level in the adjacent or the next-to-adjacent year, the missing observation is replaced with the observation from the closest available year. The procedure first fills in lacking observations with data available in $t-1$, then $t-2$. For observations which are still missing after this step, the data is imputed with available information in $t+1$, then $t+2$. This imputation procedure is based on the assumption of persistence in sectoral employment and R&D expenditure in the short run.

Alternatively, we could have interpolated missing observations by pooling the available data and running a fixed effects regression with year, country, and industry fixed effects. We choose not to impute missing observations in this way because this method puts relatively more structure on the common components of the underlying R&D process across countries and sectors.

For China, the data is taken from the Yearbook Database. For 2001-2009, the full set of information is reported for nominal R&D expenditure and full time equivalent of R&D personnel. The series corresponds to R&D activity of large and medium-sized enterprises, and includes all such enterprises whether of mixed, government, or private ownership. The series is reported in China Statistical Yearbooks on Science and Technology in 2002-2007 for 2001-2006, and in the chapter 'Education, Science, and Technology' of annual China Statistical Yearbooks in 2007-2011 for 2006-2010.

For 1995-2000, we use the data on total R&D activity in manufacturing reported in China Statistical Yearbooks on Science and Technology in 1996-2001. We compute average sectoral shares in total R&D activity in

2001-2003, and we distribute the reported totals using this weighting system. We cross-check the quality of this imputation by comparing obtained data with information on R&D activity in high-tech industries in 1995-2000 which is reported at the sectoral level in China Statistics Yearbooks on High Technology Industry (2002, 2003, 2007).

D.2 Results of two stage estimation

In this part of the appendix we report results of the two stage estimation procedure for the benchmark model in which labor and inputs are assumed to be combined in the same way across sectors. We then report first stage results of the model estimated in the core of the paper in which the share of inputs is allowed to be sector-specific.

D.2.1 Benchmark specification: no sector specifics in the production function

Table 25: First stage: Measured TFP ($\zeta^k = \zeta$)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.009***	0.008***		
	(0.003)	(0.003)		
K-stocks	0.233***	0.234***		
	(0.005)	(0.005)		
R&D: expenditure			0.090***	0.092***
			(0.002)	(0.002)
Workforce	-0.196***	-0.196***	-0.089***	-0.093***
	(0.004)	(0.004)	(0.005)	(0.004)
Cons	-1.601***	-1.751***	0.710***	0.665***
	(0.084)	(0.084)	(0.050)	(0.048)
Obs	4196	4196	4833	4833
R^2	0.384	0.398	0.167	0.177
Shea's $pcorr^2$	0.153	0.156	0.0980	0.0974
Angrist-Pischke F-stat	188.1	192.8	181.5	182.3
Angrist-Pischke χ^2	377.7	387.1	182.1	183.0

Depvar is log of measured TFP: $\overline{z}_{i,t}^k$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

As a benchmark, we estimate the model assuming that labor and inputs are combined in the same way by all sectors: $\zeta^k = \zeta$. In this case the proximity mechanism plays no role in the pattern of intersectoral specialization because there is no sectoral variation in the input component of production costs. We allow for a sector-specific component of factor costs because

Table 26: First stage: Hourly wages ($\zeta^k = \zeta$)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.083***	0.083***		
	(0.006)	(0.006)		
K-stocks	0.798***	0.807***		
	(0.009)	(0.009)		
R&D: expenditure			0.362***	0.374***
			(0.006)	(0.006)
Workforce	-0.937***	-0.949***	-0.567***	-0.589***
	(0.008)	(0.007)	(0.012)	(0.012)
Cons	-5.412***	-5.424***	1.846***	2.219***
	(0.143)	(0.145)	(0.114)	(0.109)
Obs	4196	4196	4833	4833
R^2	0.732	0.748	0.439	0.463
Shea's $pcorr^2$	0.290	0.292	0.249	0.247
Angrist-Pischke F-stat	1106	1204	552.1	573.0
Angrist-Pischke χ^2	2220	2417	554.0	575.0

Depvar is log of measured hourly wage: $\overline{\nu}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

the skill composition of the workforce remains sector-specific. Consequently, the pattern of revealed comparative advantage is determined by technology stocks and labor endowments. In the first stage, measured TFP and hourly wages are instrumented on a common set of instruments. Instruments include year fixed effects because they are included in the second stage of the estimation.

First stage results for the benchmark case are reported in tab.25 and tab.26, together with statistics on instruments' performance. Results are reported for the four specifications used throughout the paper which differ by the set of instruments. In columns (I) and (II), R&D personnel and real capital stocks are used as proxies of sectoral technology stocks, and the estimation window is 1995-2007. In columns (III) and (IV), deflated R&D expenditure is used as a proxy of technology stocks by considering that real expenditure on R&D captures variation in production costs due to technology improvement. The estimation window is 1995-2009. Specifications differ in the measure of sectoral workforce: it is efficiency-adjusted in (I) and (III), and not efficiency-adjusted in (II) and (IV).

Across specifications, Angrist-Pischke χ^2 attests that the null of underidentification of each of the endogenous regressors is rejected.[63] Shea's partial

[63] AP χ^2 is constructed by partialling out the linear projection of the other endogenous regressor. AP F-stat attests that endogenous regressors are not weakly identified.

R^2 shows that the loss of precision in the second stage linked to using instrumented endogenous regressors is relatively minor.[64]

Tab.27 reports results of the second stage. The heterogeneity parameter θ is precisely estimated at 4.4;4.5. It is little sensitive to the set of instruments used. The coefficient on hourly wages is also precisely estimated at -1.3;-1.4. The range of point estimates corresponds to a factor share of inputs in gross output of $.68 - .71$. The corresponding value of ζ for the EU-15 computed with data on gross output and expenditure on inputs in WIOD increases from .67 in 1995 to .71 in 2009, with a mean of .68 for the whole period.

Table 27: Second stage: Estimated parameters ($\zeta^k = \zeta$)

	(I)	**(II)**	**(III)**	**(IV)**
TFP	4.455***	4.386***	4.512***	4.442***
	(0.364)	(0.363)	(0.412)	(0.416)
WAGE	-1.337***	-1.281***	-1.431***	-1.362***
	(0.088)	(0.086)	(0.089)	(0.087)
Obs	4196	4196	4833	4833
Hansen J	2.118	1.578		
Hansen J p-val	0.146	0.209		
Kleibergen-Paap rk LM	385.2	386.5	302.1	298.4
Cragg Donald Wald F	240.1	246.7	221.8	221.4

2-step GMM estimation. Depvar is estimated exporter-sector dummy: $\widehat{fe}^k_{i,t}$.
Regressors are logs of instrumented TFP and wages. Wages are efficiency adjusted in (II) and (IV).
(I)-(IV) differ in the set of instruments. Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV).
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1
Year fixed effects included in each specification.

Reported values of Kleibergen-Paap rk LM and Cragg Donald Wald F statistics attest that instruments pass respectively the underidentification and weak identification tests across specifications.[65] As the equation is overidentified in the first two specifications, we report the result of the test of overidentifying restrictions (Hansen J statistic). The joint null that instruments are uncorrelated with the error term and correctly excluded from the estimation is not rejected at conventional significance levels.

[64]Loss in precision is approx. the reciprocal of the partial correlation: $1/pcorr$.

[65]The underidentification test rejects the null that the matrix of reduced form coefficients is not full rank. The value of the statistic in the weak identification test corresponds to a case in which instrumental variables estimator is not source of bias due to weak instruments.

D.2.2 First stage results with sector specific production functions

Here we report first stage results of the model estimated in the core of the paper in which the proximity mechanism is active because we allow the factor share of inputs to be sector-specific.

Table 28: First stage : Measured TFP (sector specific slopes)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.059***	0.057***		
	(0.004)	(0.004)		
K-stocks	0.235***	0.235***		
	(0.006)	(0.006)		
R&D: expenditure			0.167***	0.167***
			(0.004)	(0.004)
Workforce	-0.232***	-0.229***	-0.154***	-0.154***
	(0.004)	(0.004)	(0.005)	(0.005)
Workforce 19	-0.008***	-0.008***	-0.014***	-0.014***
Workforce 20	-0.000	-0.001	0.007***	0.007***
Workforce $21-22$	0.003*	0.002*	0.012***	0.013***
Workforce 24	-0.039***	-0.041***	-0.043***	-0.046***
Workforce 25	-0.012***	-0.013***	-0.017***	-0.018***
Workforce 26	-0.008***	-0.008***	-0.003*	-0.003*
Workforce $27-28$	-0.014***	-0.015***	-0.011***	-0.012***
Workforce 29	-0.005***	-0.005***	-0.018***	-0.020***
Workforce $30-33$	-0.016***	-0.017***	-0.031***	-0.034***
Workforce $34-35$	-0.035***	-0.037***	-0.044***	-0.047***
Workforce $36-37$	0.009***	0.009***	0.001	0.001
Obs	4196	4196	4833	4833
R^2	0.522	0.536	0.367	0.380
Shea's $pcorr^2$	0.175	0.193	0.147	0.162
Angrist-Pischke F-stat	223.1	245.3	275.2	301.4
Angrist-Pischke χ^2	449.2	493.8	276.8	303.2

Depvar is log of measured TFP: $\bar{z}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Coef. Workforce corresponds to elasticity for sector $17-18$.
For every other sector: elasticity given by sum of coef. Workforce and coef. of sector.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

To estimate this model we need instrumented sectoral hourly wages and instrumented TFP. Consequently, in the first stage we run 13 regressions in which measured TFP and hourly sectoral wages are regressed on a common set of instruments which include R&D personnel and real capital stocks in (I) and (II) (deflated R&D expenditure in (III) and (IV)) together with the workforce of each of the 12 sectors. In (I) and (III) sectoral workforce is efficiency-adjusted. In (II) and (IV) we use raw data on hourly wages and number of persons engaged in the sector.

Tab.28 reports first stage results for measured TFP in (I)-(IV). Tab.29

reports first stage results for hourly wages in the benchmark sector (textiles).[66] Across specifications the instruments pass underidentification and weak identification tests for each of the endogenous regressors.

Table 29: First stage : Hourly wages (sector specific slopes)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.131***	0.132***		
	(0.009)	(0.009)		
K-stocks	0.829***	0.836***		
	(0.010)	(0.010)		
R&D: expenditure			0.513***	0.525***
			(0.009)	(0.009)
Workforce	-1.002***	-1.010***	-0.705***	-0.723***
	(0.008)	(0.008)	(0.012)	(0.012)
Workforce 19	-0.015***	-0.016***	-0.031***	-0.034***
Workforce 20	-0.002	-0.002	0.028***	0.030***
Workforce $21-22$	-0.000	-0.001	0.038***	0.041***
Workforce 24	-0.059***	-0.064***	-0.070***	-0.077***
Workforce 25	-0.024***	-0.026***	-0.037***	-0.041***
Workforce 26	-0.038***	-0.041***	-0.015***	-0.017***
Workforce $27-28$	-0.012***	-0.013***	0.004	0.004
Workforce 29	-0.003	-0.003	-0.049***	-0.054***
Workforce $30-33$	-0.017***	-0.019***	-0.071***	-0.077***
Workforce $34-35$	-0.034***	-0.036***	-0.062***	-0.068***
Workforce $36-37$	0.011***	0.012***	-0.012***	-0.013***
Obs	4196	4196	4833	4833
R^2	0.770	0.784	0.544	0.567
Shea's $pcorr^2$	0.334	0.379	0.311	0.341
Angrist-Pischke F-stat	394.7	644.2	270.2	394.9
Angrist-Pischke χ^2	794.5	1297	271.8	397.2

Depvar is log of measured hourly wage: $\overline{\nu}_{i,t}^{k}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Coef. Workforce corresponds to elasticity for sector $17-18$.
For every other sector: elasticity given by sum of coef. Workforce and coef. of sector.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

E Addenda on the proximity characteristic

E.1 Persistence of time-varying proximity

In fig.13-16 we document the persistence of the proximity characteristic. We plot the evolution of the microfounded measure of distance to suppliers, $PROX^{-1}$, which can be interpreted as a synthetic indicator of (inverse) centrality measured in kilometers. Inverse proximity is plotted in 1995-2009

[66] Results for the other 11 sectors are qualitatively similar (not reported).

for a subsample of countries, including the least and the most distant countries in each subgroup.

Figure 13: Inverse proximity: subset of EU15

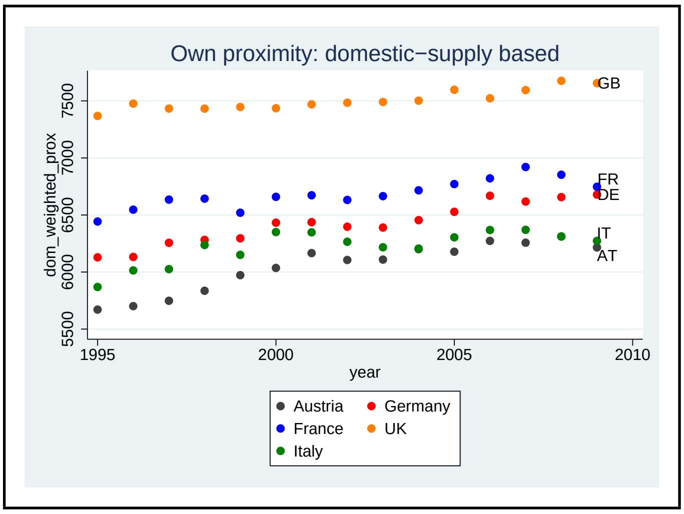


Figure 14: Inverse proximity: subset of CEECs

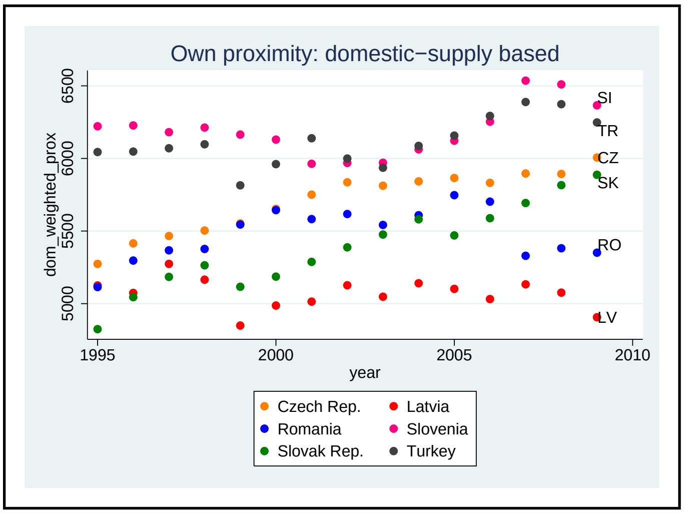


Figure 15: Inverse proximity: subset of non-European developed

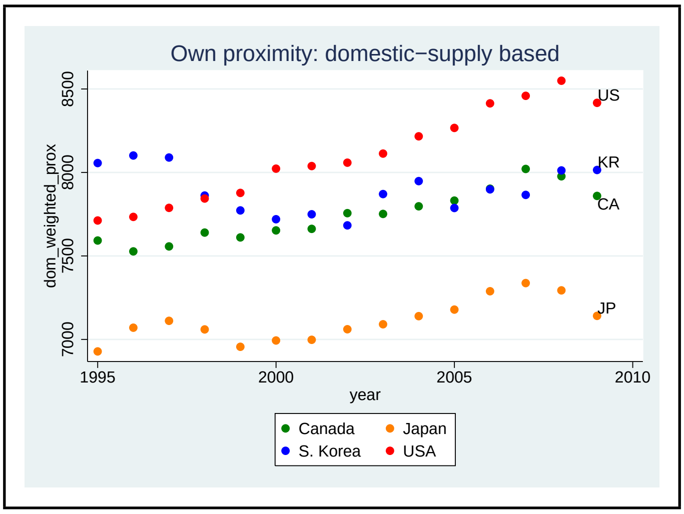
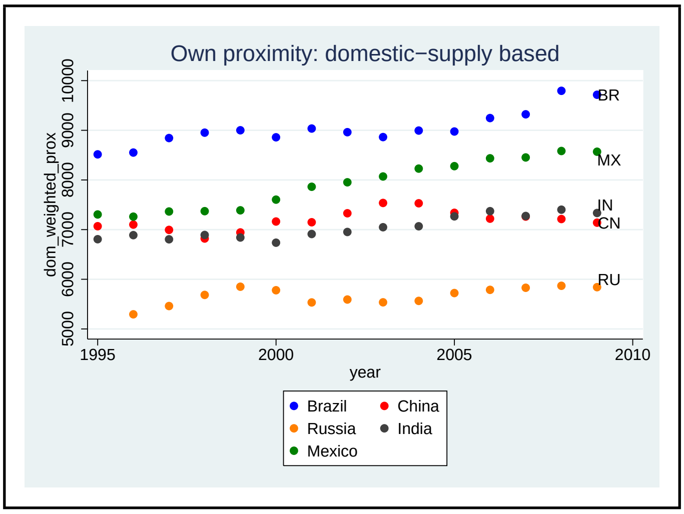


Figure 16: Inverse proximity: subset of non-European emerging



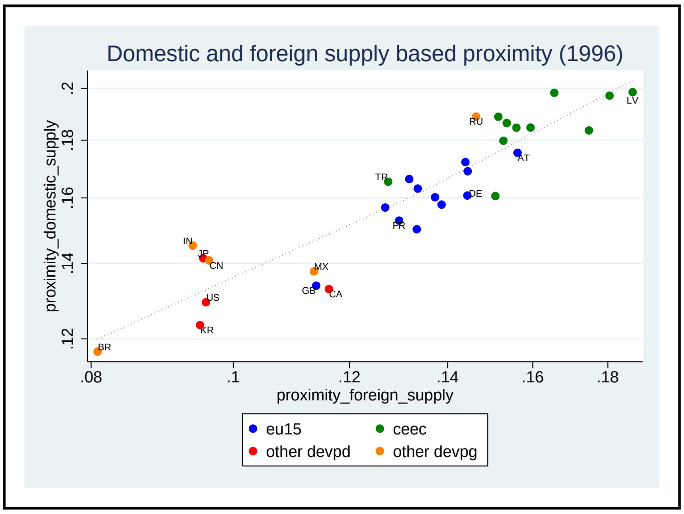
E.2 Domestic and Foreign Supply Based Proximity

The lack of output data impedes computing the proximity characteristic which we use in the core of the paper for a wider set of countries. We refer to this indicator as domestic supply based proximity (DSBP) because it weighs the components of the distance vector based on the distribution of sectoral market shares including expenditure on domestically produced varieties.

However, the availability of data on bilateral trade allows computing a proximity indicator restricted to the distribution of market shares across foreign suppliers (Foreign Supply Based Proximity) for a wider set of countries. If FSBP approximates DSBP sufficiently well, it would allow keeping more countries in the sample since FSBP can be constructed for all countries except Taiwan.

We compute this indicator for our sample, and find that FSBP picks up most of DSBP variance: 87% in 1995-2009 for the common set of countries. Fig.17-18 illustrate that the relationship becomes tighter in the recent period. Fig.19 shows the persistence of the FSBP indicator for a subset of countries.

Figure 17: DSBP and FSBP in 1996



Two factors explain differences in centrality measures obtained with these indicators: country size and differences in domestic market openness to foreign supply.

In the first case, the DSBP indicator is unbiased while the FSBP indicator underestimates proximity for relatively big countries because by construction this indicator ignores differences in domestic market size. Indeed, if trade restrictiveness is similar across destinations, then a relatively big country in terms of economic size sources a lot from itself because it is effectively least cost for a substantial chunk of varieties on the domestic market, while the smaller country sources most varieties from abroad. If internal distances are lower than external distances, then the FSBP mismeasures proximity for

Figure 18: DSBP and FSBP in 2008

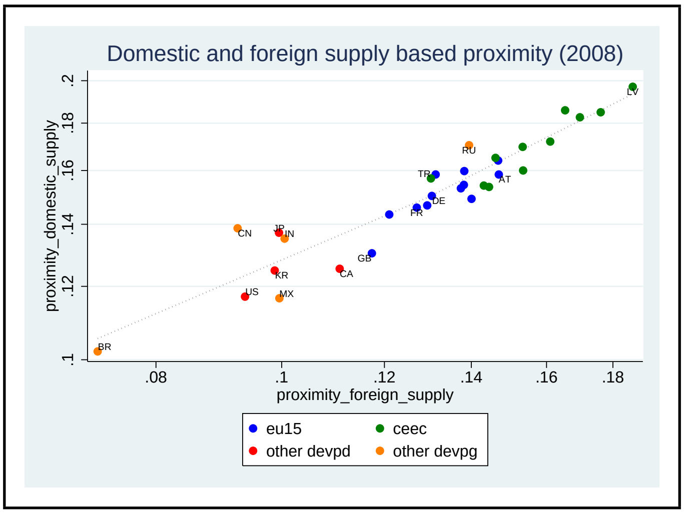
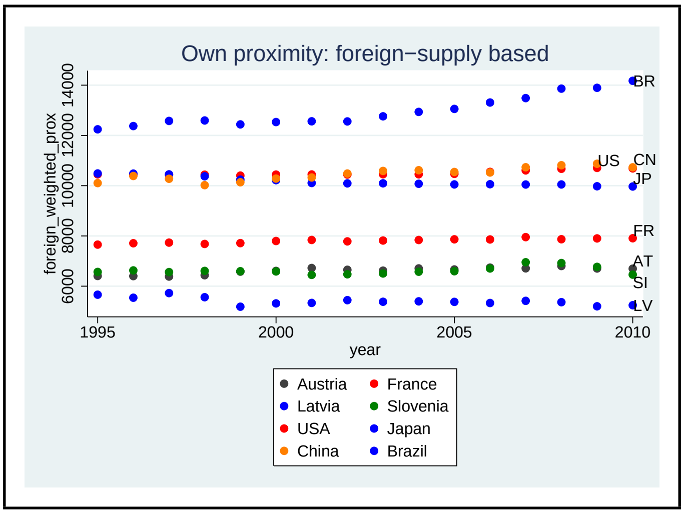


Figure 19: FSBP (subset of countries)



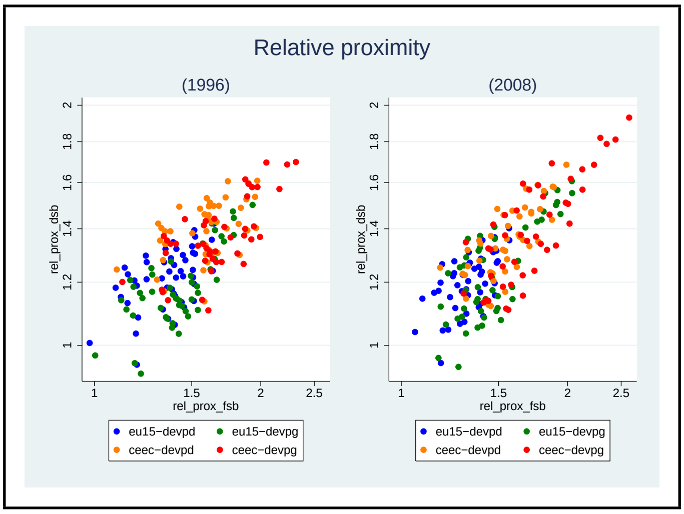
relatively big countries by putting excessive weight on bigger elements of the distance vector while being accurate for relatively small economies.

In the second case, both indicators are biased by a scalar which corresponds to trade restrictiveness of the domestic market τ_j. In this case, the FSBP indicator is an unbiased indicator of relative proximity for countries with a similar level of trade restrictiveness. However, the bias in the DSBP indicator is more severe because it also suffers from measurement error in the weighting function. Indeed, if the mark-up charged on varieties imported from abroad is relatively high, the share of varieties sourced domestically will also be relatively high, but not because the country is cost efficient in terms of fundamental productivity. Ignoring differences in trade restrictiveness leads to overestimation of domestic market share and underestimation of foreign market shares in the DSBP, and this mismeasurement of the weighting function is more severe for relatively closed markets. If internal distances are lower than external distances, the DSBP overestimates the true underlying proximity for relatively closed markets while it is relatively accurate for markets with a low overall level of trade restrictiveness (the next subsection makes this argument formally).

In our data, fig.18 illustrates that the DSBP predicted by the FSBP would be lower for several non-European economies such as Russia, China, and India while it would be higher for countries such as the USA and Canada. To explain this discrepancy, the argument of FSBP bias linked to differences in country size is not helpful because all of these countries are relatively big. Rather, this finding indicates that differences in domestic market openness to foreign supply may explain the discrepancy between DSBP and FSBP, i.e. that observed DSBP overestimates proximity for Russia, China, and India, and underestimates proximity for USA and Canada because Russia, China, and India are relatively less open to foreign supply than the USA and Canada. This line of argument is corroborated by the fact that the relationship between DSBP and FSBP becomes tighter in the recent period as illustrated in fig.20 which plots relative proximity by partner type measured with both indicators in 1996 and 2008. This evolution is consistent with the idea that initially relatively closed emerging economies are gradually opening up their domestic markets to foreign supply.

We conclude that in our data mismeasurement in the weighting function due to not correcting for destination-specific trade restrictiveness may be a bigger source of bias in the DSBP than is mismeasurement due to omitting the share of varieties sourced from the domestic market in the FSBP. Measurement error provides an additional motivation for instrumenting the DSBP indicator as we do in the core of the paper. But a feasible alternative would consist in working with the FSBP indicator instead.

Figure 20: Relative proximity: DSBP and FSBP



E.3 Proximity and trade restrictiveness of the destination

In this subsection we discuss the potential bias of domestic supply based proximity indicators constructed in the core of the paper. This bias is linked to the assumption that the trade policy component of bilateral sectoral trade costs $\tau^{k}_{ij,t}$ is fully accounted for by an exporter-specific cost component $\tau^{E,k}_{i,t}$ common across destination markets. Indeed, if destinations differ in the degree of trade restrictiveness towards foreign supply, these trade impediments will not be picked up by the exporter-specific trade cost component, and the trade cost function will be misspecified.

Waugh (2010) presents empirical evidence that the assumption of an exporter-specific trade cost component matches the data better than the alternative of trade restrictiveness common across suppliers. Nonetheless, the incidence of a destination-specific trade policy restrictiveness index $\tau_{j,t}$ needs to be considered in this paper because it would enter directly in the calculation of the proximity characteristic. Indeed, the microfounded measure of proximity is subject to measurement error in cross-section and overtime if there are substantial differences in trade restrictiveness across destinations in terms of domestic market openness to foreign supply. Ignoring this trade policy component induces a bias on relative proximity measured in cross-section. Furthermore, it may lead to misinterpreting the evolution of proximity.

As differences in trade restrictiveness across destinations are verified in the data (see below), it is necessary to control for the impact of this trade cost component on the measure of proximity to suppliers. We define the destination-specific overall trade policy restrictiveness index (OTRI) $\tau_{j,t} = 1 + t_{j,t}$ as in Kee et al. (2009) where $t_{j,t}$ is the uniform ad valorem equivalent tariff which if applied to all manufacturing imports of destination j would leave its aggregate manufacturing imports unchanged.[67]

Waugh (2010) discusses the impossibility of separately identifying exporter- and destination-specific components of trade costs using the structure of the EK model. Similarly, it is not possible to separately identify the contribution of exporter-specific trade restrictiveness (MA-OTRI) and destination-specific trade restrictiveness (OTRI) in determining bilateral trade flows using the theoretically grounded methodology developed by Kee et al. (2009). However, we do not actually compute exporter-specific trade cost components in the paper. Rather, we specify under which conditions it is possible to conduct the estimation while leaving this component of trade costs in the residual. Consequently, there is no conceptual conflict in correcting the DSBP indicator constructed in the core of the paper by considering a destination-specific trade cost component in the derivation of the proximity characteristic instead of an exporter-specific trade cost common across destinations.

First, we show that the weighting function must be adjusted to take into account the index of trade restrictiveness. Omitting time subscripts to simplify notation, sectoral expenditure is defined in the core of the paper: $X_j^k = PROD_j^k - EXP_j^k + IMP_j^k$. But the imports data is not corrected for the tariff charged on imports by the destination. Thus, the true underlying sectoral expenditure is: $\overline{X}_j^k = PROD_j^k - EXP_j^k + \tau_j IMP_j^k$. Domestic absorption of domestic varieties is still defined by: $X_{jj}^k = PROD_j^k - EXP_j^k$, but the true underlying value of bilateral imports is: $\overline{X}_{ij}^k = \tau_j X_{ij}^k$.

The true underlying domestic market share is: $\overline{\pi}_{jj}^k = X_{jj}^k / \overline{X}_j^k$ which is strictly smaller than domestic market share π_{jj}^k used in the core of the paper if sectoral imports and the uniform tariff t_j are positive: $\pi_{jj}^k = X_{jj}^k / X_j^k$, and $\overline{\pi}_{jj}^k < \pi_{jj}^k$ whenever $\overline{X}_j^k > X_j^k$. The gap with the true underlying domestic market share $\overline{\pi}_{jj}^k$ is increasing in the degree of trade restrictiveness in the destination.

Similarly, the true underlying market shares for foreign partners are higher than those used in the weighting function in the core of the paper if $X_{jj}^k > 0$ and $t_j > 0$ since $\overline{\pi}_{ij}^k = \overline{X}_{ij}^k / \overline{X}_j^k = \tau_j X_{ij}^k / \overline{X}_j^k$ and $\overline{\pi}_{ij}^k > \pi_{ij}^k$ whenever $\overline{X}_j^k < \tau_j X_j^k$.

[67] See Anderson and Neary (1996) for the theory behind welfare-based TRI.

Since $\overline{X}_j^k = X_j^k + t_j IMP_j^k$, it is obvious that the lower the degree of overall trade restrictiveness, and the closer are uncorrected market shares to the true underlying weights which should be used in the computation of proximity. We conclude that under the assumption that the internal component of the distance vector is smaller than all external components, the higher the trade restrictiveness in the destination, and the stronger is the bias in the uncorrected proximity indicator linked to giving excessive weight to the internal component and reduced weight to external components of the distance vector.

Second, we show that in constructing the proximity characteristic, the overall indicator must be rescaled by the trade restrictiveness index and domestic market share must be rescaled by the reciprocal of the trade restrictiveness index. Again, the higher the trade restrictiveness in the destination, and the stronger is the bias due to giving excessive weight to the internal component of the distance vector. Furthermore, relative proximity must be rescaled by relative overall trade restrictiveness.

Going back to the sectoral market share equation which states the probability that country i is the least cost producer for country j across the spectrum of varieties in sector k:

$$\overline{\pi}_{ij}^k = \frac{\left[\omega_i^k \tau_{ij} \tau_j / z_i^k\right]^{-\theta}}{\Phi_j^k}$$

We bring trade cost components to the left hand side of the equation and sum across all suppliers to market j in sector k including domestic consumption of domestic varieties with $\tau_j = 1$ only for domestically sourced varieties:

$$\sum_{i \neq j} \overline{\pi}_{ij}^k \left(\tau_{ij} \tau_j\right)^\theta + \overline{\pi}_{jj}^k \tau_{jj}^\theta = \frac{\sum_{n=1}^N \left[\omega_n^k / z_n^k\right]^{-\theta}}{\Phi_j^k}$$

Define $\widetilde{\Phi}^k = \sum_{n=1}^N \left[\omega_n^k / z_n^k\right]^{-\theta}$. The sectoral price distribution parameter summarizes the price distribution of best practice in the world within sector k.

Rewrite domestic expenditure: $\overline{\pi}_{jj}^k \tau_{jj}^\theta = \left\{\overline{\pi}_{jj}^k / \tau_j^\theta\right\} \left(\tau_{jj} \tau_j\right)^\theta$ to factor out OTRI in the sum on the LHS. We now write Φ_j^k using $\widetilde{\Phi}^k$:

$$\Phi_j^k = \widetilde{\Phi}^k \tau_j^{-\theta} \left\{\sum_{i \neq j} \tau_{ij}^\theta \overline{\pi}_{ij}^k + \tau_{jj}^\theta \overline{\pi}_{jj}^k / \tau_j^\theta\right\}^{-1} \tag{31}$$

As in the core of the paper, we write the sectoral price index using (31):

$$P_j^k \quad = \quad \kappa \left[\widetilde{\Phi}^k\right]^{-1/\theta} \tau_j \left\{ \sum_{i \neq j} \tau_{ij}^{\theta} \overline{\pi}_{ij}^k + \tau_{jj}^{\theta} \overline{\pi}_{jj}^k / \tau_j^{\theta} \right\}^{1/\theta} \tag{32}$$

Using (32), the cost of the input bundle in country j is:

$$P_j \quad = \quad \kappa \left\{ \prod_{k=1}^{K} \left[\widetilde{\Phi}^k\right]^{-\gamma^k/\theta} \right\} \tau_j \left\{ \prod_{k=1}^{K} \left[\sum_{i \neq j} \tau_{ij}^{\theta} \overline{\pi}_{ij}^k + \tau_{jj}^{\theta} \overline{\pi}_{jj}^k / \tau_j^{\theta} \right]^{\gamma^k/\theta} \right\} \tag{33}$$

The corrected microfounded proximity characteristic of the exporter in a world with destination-specific trade cost components is given by:

$$\left[\widetilde{PROX}_{i,t}^{M}\right]^{-1} \quad = \quad \tau_i \prod_{s=1}^{S} \left\{ \sum_{n \neq i} \tau_{ni}^{\theta} \overline{\pi}_{ni}^k + \tau_{ii}^{\theta} \overline{\pi}_{ii}^k / \tau_i^{\theta} \right\}^{\gamma^s/\theta} \tag{34}$$

Relative proximity indicators constructed in the core of the paper must be corrected by the scalar $\tau_i/\tau_{i'}$. Furthermore, domestic market share must be rescaled by $\tau_i^{-\theta}$ in computation of sectoral proximity. Finally, all bilateral trade shares must be adjusted to take into account trade restrictiveness of the destination.[68]

Table 30 shows that differences in trade restrictiveness across destinations are non negligible. For example, the uniform AVE tariff is three times higher for China relatively to EU countries, and almost seven times higher for Brazil.

Unfortunately, we do not have time-series data on OTRI in manufacturing. Data on average tariffs in the MacMAp database can be used to get additional data points.[69] Tab. 31 presents average tariffs in manufacturing in 2001 for countries of our sample in their trade with developed and developing countries. Even though this measure is not directly comparable to OTRI, the variability of average tariffs indicates that differences in trade restrictiveness across destinations are likely to be a permanent source of measurement error in domestic supply based proximity.

We use OTRI in manufacturing computed by Kee et al. (2009) in 2008 (see col.3 in tab.30) to construct proximity indicators which take into account

[68] It is straightforward to show that to correct for trade restrictiveness gaps in FSBP, it suffices to rescale by $\tau_i/\tau_{i'}$. Indeed, market shares across foreign suppliers are invariant to a destination-specific mark-up on imports.

[69] See Bouet et al. (2008).

Table 30: OTRI (t_i) in 2008 (KNO)

	OTRI	**OTRI-AGRI**	**OTRI-MANUF**	**ratio-all**	**ratio-agri**	**ratio-manuf**
EU27	**0.05**	0.37	0.03	**1.0**	1.0	1.0
Canada	**0.05**	0.19	0.04	**1.0**	0.5	1.2
USA	**0.06**	0.18	0.05	**1.1**	0.5	1.5
Switzerland	**0.05**	0.54	0.01	**1.0**	1.5	0.4
Turkey	**0.06**	0.24	0.05	**1.3**	0.7	1.9
Norway	**0.08**	0.60	0.01	**1.7**	1.6	0.3
Japan	**0.09**	0.50	0.04	**1.7**	1.4	1.3
China	**0.09**	0.15	0.09	**1.9**	0.4	3.0
Singapore	**0.13**	0.51	0.12	**2.6**	1.4	4.0
India	**0.15**	0.48	0.14	**3.0**	1.3	4.8
Russia	**0.16**	0.25	0.15	**3.2**	0.7	5.0
Mexico	**0.17**	0.29	0.15	**3.3**	0.8	5.2
Brazil	**0.20**	0.22	0.20	**4.0**	0.6	6.8
Croatia	**0.01**	0.09	0.01	**0.3**	0.2	0.2

Source: Data provided by Kee, Nicita, and Olarreaga as an .xls file.

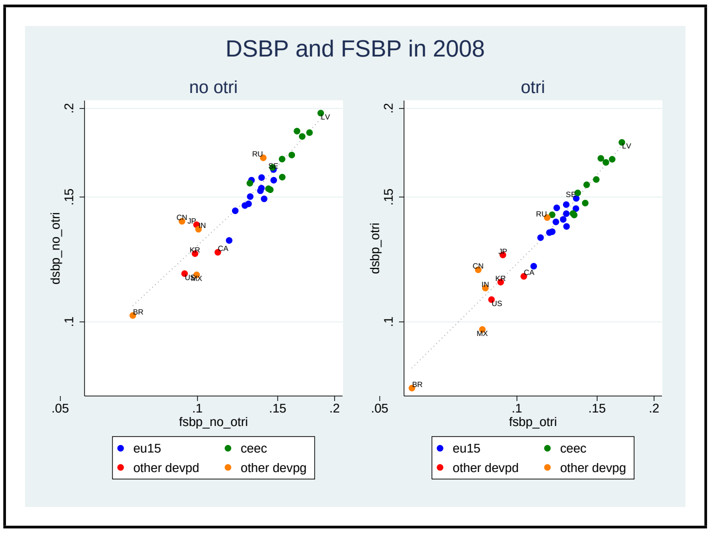
Table 31: Average tariffs in manufacturing (MacMAp database, 2001)

	Devpd t_i	**Devpg** t_i	**Ratio Devpd** t_i/t_{eu15}	**Ratio Devpg** t_i/t_{eu15}
EU15	**0.022**	**0.008**	**1.00**	**1.00**
Japan	0.006	0.003	0.29	0.38
Canada	0.020	0.006	0.89	0.77
USA	0.016	0.007	0.70	0.88
Switzerland	0.012	0.014	0.53	1.67
Korea	0.055	0.047	2.44	5.61
Estonia	0.000	0.000	0.00	0.00
Latvia	0.006	0.006	0.27	0.73
Lithuania	0.002	0.009	0.10	1.07
Turkey	0.020	0.034	0.88	4.03
Croatia	0.024	0.024	1.07	2.86
Poland	0.036	0.069	1.61	8.25
Hungary	0.040	0.042	1.77	4.99
Slovakia	0.043	0.026	1.92	3.06
Czech Republic	0.043	0.026	1.93	3.06
Bulgaria	0.047	0.045	2.09	5.41
Romania	0.076	0.085	3.36	10.14
Indonesia	0.055	0.045	2.44	5.42
Mexico	0.073	0.128	3.27	15.28
Russian Federation	0.100	0.082	4.45	9.75
Taiwan	0.102	0.060	4.55	7.19
Thailand	0.110	0.085	4.88	10.19
Malaysia	0.122	0.095	5.45	11.37
Brazil	0.126	0.084	5.61	10.01
China	0.137	0.099	6.08	11.77
India	0.311	0.292	13.84	34.84

Source: MacMAp database (2001). Data provided by Kee, Nicita, and Olarreaga as an .xls file.

differences in trade restrictiveness across destinations. Fig.21 plots OTRI-corrected DSBP and FSBP in 2008 in the RHS panel and the indicators used in the core of the paper on the LHS. The fit is slightly improved by correcting for differences in trade restrictiveness. On the RHS, FSBP explains 94% of total variance in DSBP, while it explains a slightly lower share (89%) on the LHS. This finding corroborates the intuition that measurement error in the DSBP linked to overestimating centrality for relatively closed markets partly explains the discrepancy between domestic and foreign supply based proximity indicators.

Figure 21: OTRI-correction: DSBP and FSBP



Further, the correlation of the DSBP measure with the indicator of proximity endowment is enhanced once it is corrected for differences in domestic market openness. Proximity endowment is a strong predictor of domestic supply based proximity whether or not DSBP is OTRI-corrected (see fig.22). As the proximity endowment is an instrument which captures the component of proximity independent of a specific distribution of market shares, it is likely to reduce measurement error in the DSBP indicator linked to not controlling for differences in overall trade restrictiveness of the destination.

Figure 22: OTRI-correction: DSBP and ENDOWMENT