Online Appendices for Social Networks, Ethnicity, and Entrepreneurship

William Kerr and Martin Mandorff

Empirical Appendix Tables

Table A1a: Ethnic group tabulations

Ethnic group	Self-Empl. OVER1	Self-Empl. OVER2	Self-Empl. OVER3	Self-Empl. OVER4	All workers OVER1	In-marriage rate	Self-Empl. share	All worker count	Population count
Nepal	40.7	30.0	67.6	157.5	4.2	43%	5%	4,784	8,935
Senegal	37.0	67.6	81.4	123.4	5.6	62%	19%	4,689	7,351
Zimbabwe	36.5	39.5	124.2	228.8	2.6	43%	9%	3,482	8,955
Yemen	36.3	55.2	57.6	61.6	6.5	88%	20%	6,196	10,642
Ethiopia	34.1	56.8	63.1	65.7	8.9	53%	12%	23,962	52,577
Eritrea	31.8	47.6	52.0	61.7	5.6	67%	14%	6,167	12,843
Somalia	27.6	40.8	44.9	53.8	4.3	78%	6%	9,092	18,326
Oceania, ns/nec	25.5	14.9	47.8	113.3	3.1	39%	3%	5,216	16,053
India: Gujarati	25.3	48.8	69.5	83.7	3.1	91%	19%	54,867	117,871
Sudan	25.1	28.6	42.2	48.1	3.0	73%	6%	6,607	10,671
Cape Verde	24.8	38.2	68.0	105.4	2.6	68%	7%	6,154	18,176
Uganda	24.7	43.7	55.9	67.8	2.9	56%	14%	4,357	9,620
Bermuda	24.0	33.3	49.3	89.8	2.9	30%	9%	2,081	11,238
Bangladesh	21.6	32.0	46.6	49.8	5.4	80%	11%	36,267	58,529
Sierra Leone	20.3	25.0	39.9	44.9	4.4	54%	9%	6,537	15,101
Kuwait	18.2	29.7	33.5	42.4	3.4	57%	13%	6,555	12,490
St. Kitts-Nevis	16.9	21.2	23.7	34.0	2.1	62%	4%	2,942	8,735
Saudi Arabia	15.9	24.7	41.9	50.9	2.8	37%	10%	4,022	9,725
Liberia	15.7	14.7	41.9	119.6	2.6	45%	6%	10,617	28,936
American Samoa	15.6	19.8	24.0	31.7	2.5	44%	4%	3,565	21,894
Ghana	15.5	26.7	40.9	42.3	2.8	63%	8%	27,176	51,921
South America, ns	15.1	12.2	33.2	43.4	2.5	31%	10%	3,178	11,686
Cameroon	14.5	16.5	24.4	32.1	3.1	58%	9%	5,045	9,627
Nigeria	13.6	29.9	37.2	40.2	3.9	63%	14%	53,654	102,763
Chaldean	13.1	29.2	42.1	74.2	6.1	84%	24%	12,996	31,038
Dominica	12.8	10.5	30.6	40.4	1.9	58%	6%	4,488	12,282
Bahamas	12.2	12.1	29.4	45.9	2.0	41%	8%	4,373	19,404
Tanzania	11.9	15.2	31.9	42.7	3.4	49%	20%	3,785	8,908
Haiti	11.4	22.5	34.2	36.5	2.6	72%	8%	114,602	295,424
Americas, ns	11.2	9.2	24.0	27.8	2.3	40%	11%	2,488	7,029
Partial Nordic Region	11.1	1.4	58.8	71.0	2.2	16%	16%	4,571	14,527
Singapore	10.7	11.0	19.4	28.8	2.6	39%	8%	5,316	14,470
Belgium	10.6	2.3	66.8	95.7	2.1	26%	14%	6,636	23,591
Morocco	10.4	7.5	47.6	59.9	3.4	27%	14%	14,515	29,812
Pakistan	10.3	20.2	23.2	29.7	5.1	76%	17%	85,400	144,403
Afghanistan	10.2	15.6	26.4	27.1	4.7	78%	19%	12,573	28,075
Polynesia	9.1	11.4	15.9	18.0	2.3	68%	12%	8,727	22,766
India: Punjabi	9.0	14.7	24.4	25.6	6.0	94%	21%	35,325	68,412
Cyprus	8.5	4.7	26.0	33.5	2.8	43%	21%	3,301	7,638
Africa, ns/nec	8.4	20.2	25.1	27.7	2.4	42%	12%	49,046	104,639
Antigua-Barbuda	8.2	7.8	24.0	26.4	2.0	26%	10%	4,359	13,980
Baltic States	7.9	14.7	34.0	48.0	1.7	41%	17%	7,138	27,865
Dominican Republic	7.6	15.2	21.2	26.9	2.7	60%	11%	162,086	458,705
Indochina, ns	7.6	14.5	21.5	27.7	2.4	68%	23%	21,790	43,819
Iraq	7.4	6.6	24.2	46.1	2.5	58%	18%	18,494	32,852
Jordan	7.3	13.3	20.9	21.9	3.2	68%	27%	18,945	32,794
Korea	7.0	14.1	27.5	33.7	3.9	53%	34%	185,099	574,104
Sri Lanka (Ceylon)	6.8	10.7	14.4	14.5	2.4	48%	16%	9,546	19,756
Melanesia	6.7	7.9	22.9	26.3	1.9	69%	9%	8,338	22,856
Cambodia	6.7	8.8	13.6	30.5	2.1	74%	14%	29,578	89,044
Syria	6.6	6.1	28.1	37.4	4.6	53%	29%	16,533	27,237
Former USSR	6.6	13.4	22.4	35.6	2.9	50%	17%	10,763	24,733
Bulgaria	6.5	6.6	17.1	19.8	2.1	51%	16%	11,236	24,332
Norway	6.4	4.5	19.0	29.1	2.7	24%	16%	6,307	20,178
Grenada	6.3	3.1	27.2	38.7	2.4	54%	13%	6,568	19,247
Burma (Myanmar)	6.2	9.5	20.9	24.6	2.0	60%	11%	10,730	23,431
Kenya	6.0	11.1	23.2	25.8	2.5	52%	13%	13,896	30,690
Northern Ireland	6.0	2.4	30.1	35.9	2.3	23%	17%	2,900	8,258
Greece	5.7	8.2	13.1	13.4	2.7	61%	31%	52,382	125,152
Other Caribbean	5.4	3.3	17.7	32.0	1.7	40%	11%	6,803	23,826
Europe, ns.	5.3	3.2	22.4	26.5	2.0	32%	16%	6,471	23,750
St. Vincent	5.3	2.0	14.2	18.3	2.3	36%	8%	5,404	15,237
Panama	5.3	3.5	29.4	40.3	1.4	25%	8%	19,406	93,243
Armenian	5.1	7.8	25.0	29.8	3.5	81%	28%	38,206	93,455
Denmark	4.9	1.5	20.3	23.8	2.1	12%	21%	7,153	20,760
Thailand	4.9	7.8	10.1	10.6	2.1	31%	14%	25,131	111,254
Austria	4.9	2.7	23.2	26.5	2.1	26%	22%	8,528	37,079
India: Other	4.5	7.5	10.2	12.3	3.7	67%	10%	313,091	564,570
U.S. Virgin Islands	4.5	2.0	17.5	25.1	1.6	43%	9%	8,487	34,759
New Zealand	4.4	5.0	20.4	26.1	1.9	24%	15%	7,980	19,230

Table A1a: Ethnic group tabulations

Ethnic group	Self-Empl. OVER1	Self-Empl. OVER2	Self-Empl. OVER3	Self-Empl. OVER4	All workers OVER1	In-marriage rate	Self-Empl. share	All worker count	Population count
Albania	4.3	4.7	8.5	9.1	2.4	85%	10%	11,550	24,248
China	4.2	7.8	10.1	10.5	2.9	77%	12%	289,651	653,687
Turkey	4.1	5.2	17.4	22.9	2.7	46%	19%	26,617	57,076
Barbados	4.1	2.6	13.8	15.0	2.0	46%	6%	12,694	37,166
Paraguay	4.1	3.1	12.3	15.9	2.6	42%	26%	3,144	7,843
Vietnam	4.0	6.5	11.7	12.9	2.3	73%	12%	268,627	718,423
St. Lucia	4.0	2.3	11.3	18.0	2.1	51%	12%	3,802	9,885
Belize/British Honduras	3.9	2.6	15.3	30.6	1.8	44%	9%	9,194	29,934
Lebanon	3.9	7.4	18.2	23.4	2.3	49%	25%	33,995	64,639
Northern Africa	3.8	5.7	11.9	12.5	2.5	48%	17%	48,650	87,005
Laos	3.8	2.4	15.5	22.1	2.2	80%	7%	48,346	141,048
Romania	3.6	1.3	36.5	62.0	1.4	47%	18%	34,760	86,443
Sweden	3.6	1.6	15.4	19.6	2.0	17%	17%	11,498	34,566
Croatia	3.6	3.5	15.8	27.7	1.5	50%	18%	10,851	25,577
Malaysia	3.6	4.3	8.6	9.5	2.5	35%	11%	16,103	38,138
Indonesia	3.6	3.8	12.1	16.4	1.7	24%	11%	18,160	50,484
Switzerland	3.6	3.1	16.7	23.6	2.0	21%	18%	11,996	31,231
Ecuador	3.5	5.7	14.8	16.3	2.0	39%	9%	91,506	212,967
Guam	3.5	3.0	7.9	9.9	1.7	29%	5%	7,834	43,745
Bolivia	3.4	2.3	26.9	42.3	1.9	30%	10%	13,893	37,384
Ukraine	3.4	5.1	13.5	14.5	1.6	57%	14%	59,433	143,265
Italy	3.3	4.8	9.5	9.7	1.8	57%	24%	87,551	289,037
Uruguay	3.3	1.4	13.5	14.4	1.6	24%	22%	8,297	19,269
Taiwan	3.3	5.9	8.8	9.1	2.2	46%	18%	80,135	239,620
Hong Kong and Macau	3.2	4.6	8.4	11.0	2.2	66%	14%	47,605	154,484
Brazil	3.2	4.9	23.3	25.9	1.6	36%	15%	59,408	154,828
Guyana/British Guiana	3.2	4.3	12.9	16.1	2.1	59%	9%	55,565	151,927
USSR/Russia	3.0	5.3	12.7	13.8	1.9	61%	15%	97,769	249,585
France	2.9	2.0	20.8	24.1	1.8	29%	14%	36,805	130,567
Iran	2.8	2.3	11.1	16.1	1.9	55%	28%	85,202	178,670
Venezuela	2.8	2.4	10.7	13.8	1.4	33%	14%	27,603	76,541
Philippines	2.7	4.3	10.1	13.2	1.9	50%	7%	304,598	1,027,398
Japan	2.6	3.2	6.5	7.7	1.9	31%	14%	79,389	303,281
South Africa (Union of)	2.6	2.9	10.0	15.1	2.0	34%	20%	19,762	47,921
Israel/Palestine	2.5	2.2	12.9	19.6	1.9	52%	30%	35,990	82,664
Trinidad and Tobago	2.5	2.5	9.7	13.1	1.5	44%	10%	46,055	141,913
Netherlands	2.4	1.7	7.9	8.9	1.8	23%	19%	20,333	64,956
Portugal	2.4	2.9	11.1	13.3	1.8	62%	15%	47,004	149,179
Costa Rica	2.4	1.6	9.5	14.5	1.5	25%	9%	19,433	52,472
Australia	2.4	1.8	8.1	8.5	2.1	27%	17%	16,336	48,237
Hungary	2.4	2.0	16.5	22.6	1.4	30%	24%	18,848	51,325
Yugoslavia	2.2	1.8	7.5	8.5	1.5	52%	9%	57,896	131,241
Guatemala	2.1	2.1	8.7	10.6	1.8	43%	8%	162,886	358,480
Peru	2.1	1.1	11.7	15.5	1.4	23%	12%	83,560	204,158
Argentina	2.1	1.9	9.2	11.5	1.4	27%	21%	35,789	91,664
Chile	2.0	1.2	7.4	8.2	1.4	19%	17%	23,556	58,260
Czech	2.0	1.5	15.7	20.8	1.3	31%	19%	17,862	50,681
Colombia	2.0	2.4	7.8	8.2	1.5	38%	13%	131,514	365,985
Ireland	2.0	2.3	6.2	6.8	1.4	43%	20%	41,981	100,409
Nicaragua	1.9	1.7	10.9	17.3	1.4	36%	11%	54,305	162,528
Honduras	1.9	2.1	4.7	5.4	1.7	42%	7%	87,059	210,264
El Salvador	1.9	2.0	4.1	7.2	1.7	54%	8%	260,256	630,779
Mexico	1.8	2.4	4.7	4.9	1.9	83%	8%	2,764,037	6,335,953
United Kingdom	1.7	1.9	4.8	5.0	1.5	26%	16%	157,918	519,789
Jamaica	1.7	2.1	7.2	7.9	1.5	50%	11%	124,948	409,092
Puerto Rico	1.7	1.5	6.9	14.0	1.3	62%	7%	217,852	807,876
Spain	1.7	1.3	6.1	11.5	1.5	32%	17%	22,939	77,558
Poland	1.6	1.8	5.2	5.5	1.5	55%	15%	117,444	307,017
Germany	1.6	1.0	9.0	14.8	1.3	33%	17%	98,598	725,051
Canada	1.5	1.7	3.9	5.6	1.4	25%	17%	151,273	567,555
Cuba	1.4	1.3	4.2	4.4	1.3	56%	17%	196,375	566,413

Notes: See Table 1.

Table A1b: Ethnic group tabulations

Ethnic group	Industry with most self-employed workers	Industry of max overage for self-employed	Industry with most workers
Nepal	Retail trade, n.s.	Textile mill products	Educational institutions
Senegal	Taxicab service	Apparel, fabrics, and notions	Eating and drinking places
Zimbabwe	Physicians & health practitioners	Leather and leather products	Educational institutions
Yemen	Grocery stores	Grocery stores	Grocery stores
Ethiopia	Taxicab service	Taxicab service	Taxicab service
Eritrea	Taxicab service	Taxicab service	Taxicab service
Somalia	Taxicab service	Taxicab service	Taxicab service
Oceania, ns/nec	Retail trade, n.s.	Petroleum products	Eating and drinking places
India: Gujarati	Hotels and motels	Hotels and motels	Hotels and motels
Sudan	Taxicab service	Misc. merchandise stores	Grocery stores
Cape Verde	Construction	Elementary and secondary schools	Construction
Uganda	Hotels and motels	Drug stores	Hospitals
Bermuda	Construction	Food stores, n.e.c.	Construction
Bangladesh	Taxicab service	Taxicab service	Eating and drinking places
Sierra Leone	Taxicab service	Taxicab service	Nursing and personal care facilities
Kuwait	Grocery stores	Gasoline service stations	Grocery stores
St. Kitts-Nevis	Taxicab service	Taxicab service	Construction
Saudi Arabia	Grocery stores	Jewelry stores	Educational institutions
Liberia	Taxicab service	Residential care facilities	Educational institutions
American Samoa	Services to dwellings	Taxicab service	Construction
Ghana	Taxicab service	Taxicab service	Hospitals
South America, ns	Construction	Professional/photographic equipment	Construction
Cameroon	Taxicab service	Drug stores	Hospitals
Nigeria	Taxicab service	Taxicab service	Hospitals
Chaldean	Grocery stores	Liquor stores	Grocery stores
Dominica	Construction	Direct selling establishments	Construction
Bahamas	Construction	Hospitals	Construction
Tanzania	Retail trade, n.s.	Liquor stores	Educational institutions
Haiti	Taxicab service	Taxicab service	Eating and drinking places
Americas, ns	Construction	Electrical repair shops	Construction
Partial Nordic Region	Construction	Paper and allied products	Construction
Singapore	Business services, n.e.c.	Food stores, n.e.c.	Educational institutions
Belgium	Legal services	Residential care facilities	Educational institutions
Morocco	Construction	Bus service and urban transit	Eating and drinking places
Pakistan	Taxicab service	Taxicab service	Taxicab service
Afghanistan	Taxicab service	Taxicab service	Eating and drinking places
Polynesia	Landscaping	Communications	Construction
India: Punjabi	Taxicab service	Taxicab service	Taxicab service
Cyprus	Eating and drinking places	Book and stationery stores	Eating and drinking places
Africa, ns/nec	Taxicab service	Taxicab service	Eating and drinking places
Antigua-Barbuda	Construction	Professional/photographic equipment	Construction
Baltic States	Construction	Personnel supply services	Construction
Dominican Republic	Taxicab service	Taxicab service	Construction
Indochina, ns	Grocery stores	Shoe stores	Grocery stores
Iraq	Grocery stores	Liquor stores	Construction
Jordan	Grocery stores	Grocery stores	Grocery stores
Korea	Laundry, cleaning, and garment services	Laundry, cleaning, and garment services	Eating and drinking places
Sri Lanka (Ceylon)	Physicians & health practitioners	Physicians & health practitioners	Educational institutions
Melanesia	Landscaping	Hotels and motels	Construction
Cambodia	Eating and drinking places	Retail bakeries	Eating and drinking places
Syria	Physicians & health practitioners	Liquor stores	Hospitals
Former USSR	Construction	Jewelry stores	Construction
Bulgaria	Construction	Elementary and secondary schools	Educational institutions
Norway	Construction	Child care services	Construction
Grenada	Construction	Misc. merchandise stores	Construction
Burma (Myanmar)	Eating and drinking places	Professional/photographic equipment	Eating and drinking places
Kenya	Physicians & health practitioners	Hotels and motels	Educational institutions
Northern Ireland	Construction	Elementary and secondary schools	Construction
Greece	Eating and drinking places	Eating and drinking places	Eating and drinking places
Other Caribbean	Construction	Educational institutions	Construction
Europe, ns.	Construction	Social services, n.e.c.	Construction
St. Vincent	Construction	Laundry, cleaning, and garment services	Construction
Panama	Construction	Shoe stores	Construction
Armenian	Construction	Leather and leather products	Construction
Denmark	Construction	Paper and allied products	Construction
Thailand	Eating and drinking places	Eating and drinking places	Eating and drinking places
Austria	Construction	Apparel and other finished textile products	Construction
India: Other	Physicians & health practitioners	Hotels and motels	Computer and data processing services
U.S. Virgin Islands	Construction	Apparel and accessory stores, except shoe	Construction

Table A1b: Ethnic group tabulations

Ethnic group	Industry with most self-employed workers	Industry of max overage for self-employed	Industry with most workers
New Zealand	Construction	Transportation equipment	Construction
Albania	Construction	Food stores, n.e.c.	Eating and drinking places
China	Eating and drinking places	Eating and drinking places	Eating and drinking places
Turkey	Construction	Leather and leather products	Educational institutions
Barbados	Construction	Educational services	Construction
Paraguay	Construction	Miscellaneous vehicle dealers	Construction
Vietnam	Misc. personal services	Misc. personal services	Electrical machinery and equipment
St. Lucia	Construction	Catalog and mail order houses	Construction
Belize/British Honduras	Construction	Banking	Construction
Lebanon	Construction	Gasoline service stations	Eating and drinking places
Northern Africa	Eating and drinking places	Taxicab service	Eating and drinking places
Laos	Agricultural production, crops	Textile mill products	Machinery and computing equipment
Romania	Construction	Residential care facilities	Construction
Sweden	Construction	Furniture and home furnishings	Construction
Croatia	Construction	Residential care facilities	Construction
Malaysia	Eating and drinking places	R&D and testing services	Eating and drinking places
Indonesia	Eating and drinking places	Furniture and home furnishings	Eating and drinking places
Switzerland	Misc. professional services	Farm supplies	Educational institutions
Ecuador	Construction	Taxicab service	Construction
Guam	Construction	Misc entertainment and recreation services	Construction
Bolivia	Construction	Residential care facilities	Construction
Ukraine	Construction	Taxicab service	Construction
Italy	Construction	Barber shops	Construction
Uruguay	Construction	Gasoline service stations	Construction
Taiwan	Eating and drinking places	Wholesale trade, n.s.	Electrical machinery and equipment
Hong Kong and Macau	Eating and drinking places	Textile mill products	Eating and drinking places
Brazil	Construction	Private households	Construction
Guyana/British Guiana	Construction	Metals and minerals, except petroleum	Construction
USSR/Russia	Construction	Taxicab service	Construction
France	Construction	Retail bakeries	Eating and drinking places
Iran	Construction	Apparel, fabrics, and notions	Eating and drinking places
Venezuela	Construction	Residential care facilities	Construction
Philippines	Physicians & health practitioners	Nursing and personal care facilities	Hospitals
Japan	Eating and drinking places	Museums, art galleries, and zoos	Eating and drinking places
South Africa (Union of)	Physicians & health practitioners	Metals and minerals, except petroleum	Computer and data processing services
Israel/Palestine	Construction	Sewing, needlework, and piece goods stores	Construction
Trinidad and Tobago	Construction	Child care services	Construction
Netherlands	Construction	Drugs, chemicals, and allied products	Educational institutions
Portugal	Construction	Retail bakeries	Construction
Costa Rica	Construction	Leather and leather products	Construction
Australia	Construction	Theaters and video rental	Educational institutions
Hungary	Construction	Drugs, chemicals, and allied products	Construction
Yugoslavia	Construction	Furniture and home furnishings	Construction
Guatemala	Construction	Private households	Construction
Peru	Construction	Nursing and personal care facilities	Construction
Argentina	Construction	Child care services	Construction
Chile	Construction	Chemicals and allied products	Construction
Czech	Construction	Shoe stores	Construction
Colombia	Construction	Taxicab service	Construction
Ireland	Construction	Farm-product raw materials	Construction
Nicaragua	Construction	Bus service and urban transit	Construction
Honduras	Construction	Private households	Construction
El Salvador	Construction	Private households	Construction
Mexico	Construction	Landscaping	Construction
United Kingdom	Construction	Educational institutions	Construction
Jamaica	Construction	Misc. merchandise stores	Construction
Puerto Rico	Construction	Paper and allied products	Construction
Spain	Construction	Alcoholic beverages	Construction
Poland	Construction	Museums, art galleries, and zoos	Construction
Germany	Construction	Museums, art galleries, and zoos	Construction
Canada	Construction	Furniture and home furnishings	Construction
Cuba	Construction	Shoe stores	Construction

Notes: See Table 1.

Table A2a: Pairwise correlations of various overage metrics

	Sample	Metric	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Self-employed	Weighted average overage ratio across all industries [OVER1]	1							
(2)		Weighted average overage ratio in three largest industries [OVER2]	0.925	1						
(3)		Average of three largest overage ratios for ethnic group [OVER3]	0.887	0.802	1					
(4)		Largest overage ratio for ethnic group [OVER4]	0.826	0.668	0.920	1				
(5)	All workers	Weighted average overage ratio across all industries [OVER1]	0.693	0.749	0.600	0.467	1			
(6)		Weighted average overage ratio in three largest industries [OVER2]	0.547	0.656	0.456	0.305	0.889	1		
(7)		Average of three largest overage ratios for ethnic group [OVER3]	0.568	0.652	0.539	0.418	0.897	0.801	1	
(8)		Largest overage ratio for ethnic group [OVER4]	0.515	0.602	0.501	0.391	0.833	0.724	0.946	1

Notes: Table displays correlations between ethnic group overage measures calculated on both self-employment and industry total employment. All correlations are significant at a 5% level.

Table A2b: Pairwise rank correlations of various overage metrics

	Sample	Metric	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Self-employed	Weighted average overage ratio across all industries [OVER1]	1							
(2)		Weighted average overage ratio in three largest industries [OVER2]	0.865	1						
(3)		Average of three largest overage ratios for ethnic group [OVER3]	0.894	0.707	1					
(4)		Largest overage ratio for ethnic group [OVER4]	0.858	0.647	0.969	1				
(5)	All workers	Weighted average overage ratio across all industries [OVER1]	0.817	0.809	0.667	0.601	1			
(6)		Weighted average overage ratio in three largest industries [OVER2]	0.509	0.583	0.375	0.282	0.762	1		
(7)		Average of three largest overage ratios for ethnic group [OVER3]	0.739	0.768	0.665	0.619	0.849	0.521	1	
(8)		Largest overage ratio for ethnic group [OVER4]	0.696	0.738	0.628	0.587	0.802	0.476	0.968	1

Notes: See Table A2a. Table displays rank correlations between ethnic group overage measures calculated on both self-employment and industry total employment. All correlations are significant at a 5% level.

Table A3a: OLS estimations of overage metric designs and non-parametric forms without controls

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
(0,1) Indicator: ethnic size in smallest third x	1.758+++	1.458+++	1.377+++	1.225+++
(0,1) Indicator: ethnic isolation in highest third	(0.451)	(0.447)	(0.404)	(0.443)
(0,1) Indicator: ethnic size in smallest third x	1.178+++	0.693+++	0.943+++	0.899+++
(0,1) Indicator: ethnic isolation in middle third	(0.220)	(0.252)	(0.209)	(0.249)
(0,1) Indicator: ethnic size in smallest third x	1.132+++	0.728++	1.509+++	1.378+++
(0,1) Indicator: ethnic isolation in lowest third	(0.296)	(0.348)	(0.336)	(0.326)
(0,1) Indicator: ethnic size in middle third x	0.844+++	0.946+++	1.011+++	0.829+++
(0,1) Indicator: ethnic isolation in highest third	(0.211)	(0.281)	(0.249)	(0.195)
(0,1) Indicator: ethnic size in middle third x	0.429+	0.345	0.735+++	0.787+++
(0,1) Indicator: ethnic isolation in middle third	(0.244)	(0.307)	(0.250)	(0.245)
(0,1) Indicator: ethnic size in middle third x	0.039	-0.170	0.208	0.205
(0,1) Indicator: ethnic isolation in lowest third	(0.111)	(0.173)	(0.211)	(0.151)
(0,1) Indicator: ethnic size in largest third x	0.289+	0.396	0.321	0.228
(0,1) Indicator: ethnic isolation in highest third	(0.167)	(0.258)	(0.241)	(0.165)
(0,1) Indicator: ethnic size in largest third x	-0.073	-0.107	-0.094	-0.077
(0,1) Indicator: ethnic isolation in middle third	(0.096)	(0.174)	(0.166)	(0.112)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in lowest third	Excluded group			
Adjusted R-Squared value	0.302	0.182	0.265	0.224

Notes: See Tables 1 and 3. Effects are measured relative to largest and least isolated ethnic groups.

Table A3b: OLS estimations of overage metric designs and non-parametric forms with controls

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in highest third	1.638+++ (0.399)	1.299+++ (0.407)	1.343+++ (0.401)	1.194++ (0.463)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in middle third	0.943+++ (0.184)	0.424+ (0.226)	0.795+++ (0.212)	0.773+++ (0.226)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in lowest third	1.079+++ (0.259)	0.682++ (0.296)	1.488+++ (0.317)	1.350+++ (0.316)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in highest third	0.859+++ (0.241)	0.847+++ (0.309)	1.002+++ (0.282)	0.867+++ (0.272)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in middle third	0.414+ (0.211)	0.292 (0.269)	0.711+++ (0.231)	0.777+++ (0.233)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in lowest third	-0.049 (0.096)	-0.236 (0.145)	0.154 (0.191)	0.134 (0.140)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in highest third	0.683+++ (0.212)	0.726+++ (0.271)	0.670++ (0.259)	0.561++ (0.231)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in middle third	0.163 (0.160)	0.108 (0.206)	0.153 (0.197)	0.128 (0.164)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in lowest third	Excluded group			
Adjusted R-Squared value	0.419	0.303	0.335	0.269

Notes: See Tables 1 and 3. Effects are measured relative to largest and least isolated ethnic groups. Estimations include controls for ethnic group traits.

Table A4: IV estimations with in-marriage rates in Spain 2011

	Instrumenting with predicted ethnic group size from gravity model and in-marriage rates in Spain 2011			Instrumenting with predicted ethnic group size from gravity model and in-marriage rates in UK 1991 and Spain 2011		
	First stage for size	First stage for isolation	Second stage	First stage for size	First stage for isolation	Second stage
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Instrument for size	0.624+++	0.008		0.633+++	-0.087	
	(0.067)	(0.119)		(0.075)	(0.125)	
Instrument for isolation Spain	-0.092	0.312+++		-0.059	0.135	
	(0.067)	(0.103)		(0.068)	(0.083)	
Instrument for isolation UK				-0.114+	0.491+++	
				(0.064)	(0.088)	
F-Statistic	50.6	5.0		27.1	26.4	
Inverse of log ethnic group size			0.478+++			0.456+++
			(0.153)			(0.132)
Isolation of ethnic group			0.583+++			0.350+++
			(0.219)			(0.082)
Exogeneity test p-value			0.044			0.058
Overid test p-value						0.199
			B. Including controls for ethnic group's traits			
Instrument for size	0.519+++	-0.060		0.500+++	-0.073	
	(0.094)	(0.100)		(0.098)	(0.094)	
Instrument for isolation Spain	-0.035	0.164+		-0.023	0.051	
	(0.065)	(0.096)		(0.062)	(0.083)	
Instrument for isolation UK				-0.071	0.344+++	
				(0.073)	(0.063)	
F-Statistic	15.4	1.7		9.0	11.7	
Inverse of log ethnic group size			0.407+			0.296++
			(0.238)			(0.138)
Isolation of ethnic group			1.212+			0.525+++
			(0.638)			(0.145)
Exogeneity test p-value			0.014			0.074
Overid test p-value						0.110

Notes: See Table 6. Regressions with the Spain 2011 and dual instruments have 130 and 129 observations, respectively. Regressions cluster standard errors by the ethnic groups in the Spain 2011 dataset.

Table A5a: Robustness checks on IV estimations, UK IV only

	Baseline estimation	Without sample weights	Without winsorization	Using bootstrapped standard errors	Isolation IV only with expected overage control	Dual IV with expected overage control
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.459+++	0.480+++	0.443+++	0.480+++	0.384+++	0.082
(small groups have larger values)	(0.130)	(0.138)	(0.116)	(0.142)	(0.123)	(0.220)
Isolation of ethnic group	0.316++	0.308++	0.311++	0.308	0.298+	0.252
	(0.125)	(0.135)	(0.121)	(0.201)	(0.153)	(0.187)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.294++	0.295+	0.298++	0.295	0.297+++	-0.115
(small groups have larger values)	(0.143)	(0.157)	(0.124)	(0.279)	(0.106)	(0.224)
Isolation of ethnic group	0.484+++	0.456++	0.495+++	0.456	0.469+++	0.348++
	(0.153)	(0.180)	(0.159)	(0.354)	(0.130)	(0.170)

Notes: See Table 6.

Table A5b: Robustness checks on IV estimations, combined UK and Spain IV

	Baseline estimation	Without sample weights	Without winsorization	Using bootstrapped standard errors	Isolation IV only with expected overage control	Dual IV with expected overage control
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Inverse of log ethnic group size	0.469+++	0.487+++	0.454+++	0.487+++	0.449+++	0.062
(small groups have larger values)	(0.140)	(0.144)	(0.125)	(0.163)	(0.143)	(0.229)
Isolation of ethnic group	0.419+++	0.421+++	0.411+++	0.421+++	0.543+++	0.476+++
	(0.119)	(0.126)	(0.104)	(0.133)	(0.127)	(0.163)
			B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.342++	0.344+	0.347++	0.344	0.343++	-0.082
(small groups have larger values)	(0.174)	(0.186)	(0.156)	(0.525)	(0.134)	(0.195)
Isolation of ethnic group	0.728+++	0.714+++	0.717+++	0.714	0.813+++	0.668+++
	(0.209)	(0.214)	(0.182)	(1.315)	(0.237)	(0.248)

Notes: See Table 6.

Table A6a: Robustness checks on IV estimations, UK IV only

	Baseline estimation with OVER1	Using total worker sample	Excluding natives from denominator shares	Imposing min counts on ethnic industry presence	Excluding new arrivals over the prior five years	Excluding the taxi industry
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Inverse of log ethnic group size	0.459+++	0.276+	0.606+++	0.582+++	0.323++	0.404+++
(small groups have larger values)	(0.130)	(0.164)	(0.154)	(0.156)	(0.152)	(0.137)
Isolation of ethnic group	0.316++	0.612+++	0.386+	0.347	0.535+++	0.497++
	(0.125)	(0.213)	(0.207)	(0.218)	(0.204)	(0.202)
			B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.294++	0.043	0.381+++	0.373++	0.162	0.382+++
(small groups have larger values)	(0.143)	(0.153)	(0.144)	(0.145)	(0.139)	(0.138)
Isolation of ethnic group	0.484+++	0.700+++	0.373++	0.355++	0.712+++	0.613+++
	(0.153)	(0.146)	(0.152)	(0.176)	(0.158)	(0.189)

Notes: See Table 6.

Table A6b: Robustness checks on IV estimations, combined UK and Spain IV

	Baseline estimation with OVER1	Using total worker sample	Excluding natives from denominator shares	Imposing min counts on ethnic industry presence	Excluding new arrivals over the prior five years	Excluding the taxi industry
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.469+++	0.297+	0.613+++	0.593+++	0.344++	0.417+++
(small groups have larger values)	(0.140)	(0.178)	(0.160)	(0.164)	(0.165)	(0.141)
Isolation of ethnic group	0.419+++	0.798+++	0.462++	0.464++	0.718+++	0.618+++
	(0.119)	(0.198)	(0.210)	(0.217)	(0.181)	(0.179)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.342++	0.086	0.386++	0.390++	0.209	0.380++
(small groups have larger values)	(0.174)	(0.157)	(0.159)	(0.156)	(0.156)	(0.151)
Isolation of ethnic group	0.728+++	0.920+++	0.401++	0.446++	0.952+++	0.602+++
	(0.209)	(0.212)	(0.203)	(0.219)	(0.283)	(0.220)

Notes: See Table 6.

Table A7a: IV estimations of overage metric designs, UK IV only

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
	A. Baseline estimation without controls			
Inverse of log ethnic group size	0.459+++	0.331++	0.392+++	0.394+++
(small groups have larger values)	(0.130)	(0.149)	(0.145)	(0.143)
Isolation of ethnic group	0.316++	0.483+++	0.179	0.093
	(0.125)	(0.145)	(0.163)	(0.127)
Exogeneity test p-value	0.140	0.145	0.122	0.276
	B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.294++	0.125	0.160	0.242
(small groups have larger values)	(0.143)	(0.141)	(0.175)	(0.175)
Isolation of ethnic group	0.484+++	0.645+++	0.215	0.174
	(0.153)	(0.134)	(0.224)	(0.206)
Exogeneity test p-value	0.237	0.141	0.082	0.270

Notes: See Table 6.

Table A7b: IV estimations of overage metric designs, combined UK and Spain IV

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
		A. Baseline estimation without controls		
Inverse of log ethnic group size	0.469+++	0.349++	0.411+++	0.409+++
(small groups have larger values)	(0.140)	(0.161)	(0.151)	(0.147)
Isolation of ethnic group	0.419+++	0.639+++	0.379++	0.256+
	(0.119)	(0.140)	(0.176)	(0.137)
Exogeneity test p-value	0.023	0.000	0.007	0.023
		B. Including controls for ethnic group's traits		
Inverse of log ethnic group size	0.342++	0.182	0.232	0.306+
(small groups have larger values)	(0.174)	(0.172)	(0.174)	(0.182)
Isolation of ethnic group	0.728+++	0.933+++	0.588++	0.501+++
	(0.209)	(0.231)	(0.271)	(0.193)
Exogeneity test p-value	0.006	0.002	0.025	0.055

Notes: See Table 6.

Table A8a: IV results with alternative gravity model designs for predicted size, UK IV only

	Baseline estimation	Including border in the gravity model	Including distance squared in the gravity model	Using distance and population as instruments	Using distance, population, and border as instruments	Using distance, population, and distance squared as instruments
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Inverse of log ethnic group size (small groups have larger values)	0.459+++ (0.130)	0.451+++ (0.125)	0.465+++ (0.132)	0.480+++ (0.135)	0.481+++ (0.136)	0.508+++ (0.139)
Isolation of ethnic group	0.316++ (0.125)	0.314++ (0.126)	0.316++ (0.125)	0.379+++ (0.113)	0.390+++ (0.101)	0.420+++ (0.100)
			B. Including controls for ethnic group's traits			
Inverse of log ethnic group size (small groups have larger values)	0.294++ (0.143)	0.297++ (0.136)	0.298++ (0.143)	0.343++ (0.159)	0.344++ (0.158)	0.350++ (0.158)
Isolation of ethnic group	0.484+++ (0.153)	0.484+++ (0.156)	0.485+++ (0.152)	0.644+++ (0.145)	0.638+++ (0.143)	0.665+++ (0.135)

Notes: See Table 6.

Table A8b: IV results with alternative gravity model designs for predicted size, combined UK and Spain IV

	Baseline estimation	Including border in the gravity model	Including distance squared in the gravity model	Using distance and population as instruments	Using distance, population, and border as instruments	Using distance, population, and distance squared as instruments
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Inverse of log ethnic group size	0.469+++	0.463+++	0.475+++	0.488+++	0.489+++	0.517+++
(small groups have larger values)	(0.140)	(0.138)	(0.140)	(0.143)	(0.144)	(0.145)
Isolation of ethnic group	0.419+++	0.416+++	0.422+++	0.428+++	0.439+++	0.501+++
	(0.119)	(0.120)	(0.118)	(0.120)	(0.117)	(0.118)
			B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.342++	0.344++	0.343++	0.384++	0.385++	0.389++
(small groups have larger values)	(0.174)	(0.171)	(0.172)	(0.186)	(0.186)	(0.180)
Isolation of ethnic group	0.728+++	0.729+++	0.728+++	0.829+++	0.824+++	0.847+++
	(0.209)	(0.208)	(0.208)	(0.252)	(0.255)	(0.229)

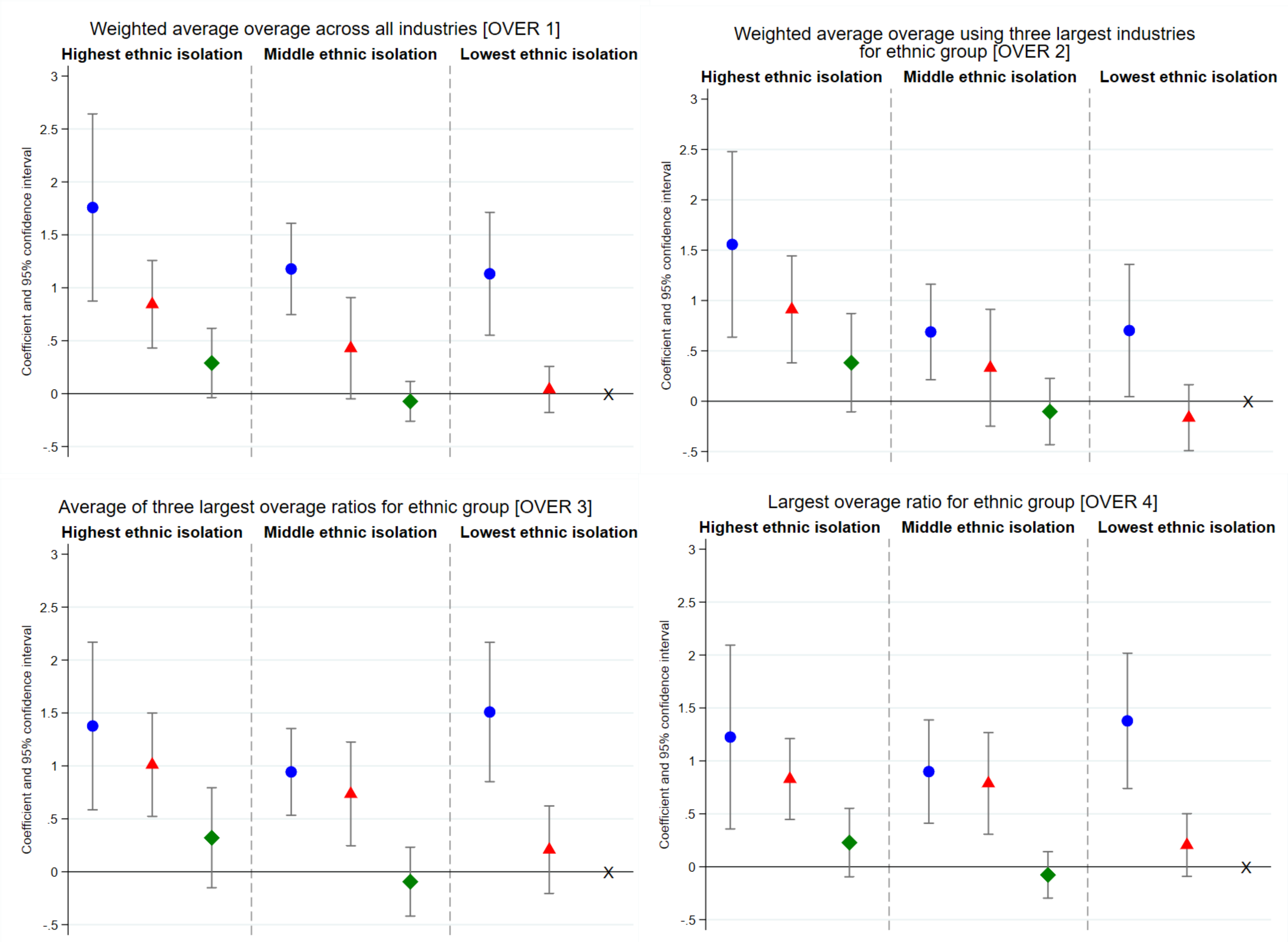
Notes: See Table 6.

Table A9: OLS estimations of individual incomes and group concentration

	Log yearly income in 2000					
	(1)	(2)	(3)	(4)	(5)	(6)
(0,1) Self employed worker	0.049+++	0.014	0.015	0.027	0.024	0.020
	(0.013)	(0.013)	(0.013)	(0.020)	(0.020)	(0.019)
Percentage of individual's group who are self-employed in the industry	0.057+++ (0.013)	0.027++ (0.012)	0.020 (0.013)			
x (0,1) Self employed worker	0.001	-0.003	0.000		0.004	-0.027
	(0.012)	(0.010)	(0.013)		(0.014)	(0.017)
Percentage of individual's group who are self-employed		0.059+++ (0.007)	0.062+++ (0.007)			
x (0,1) Self employed worker		0.042+++	0.042+++			0.080+++
		(0.008)	(0.010)			(0.015)
Percentage of individual's group who are working in the industry			0.023+ (0.012)			
x (0,1) Self employed worker			-0.000			0.029
			(0.015)			(0.021)
Person-level Traits FE	Yes	Yes	Yes	Yes	Yes	Yes
MSA-Industry FE	Yes	Yes	Yes			
MSA-Industry-Ethnicity FE				Yes	Yes	Yes
Adjusted R-Squared Value	0.246	0.248	0.248	0.308	0.308	0.308
Observations	404,467	404,467	404,467	404,467	404,467	404,467

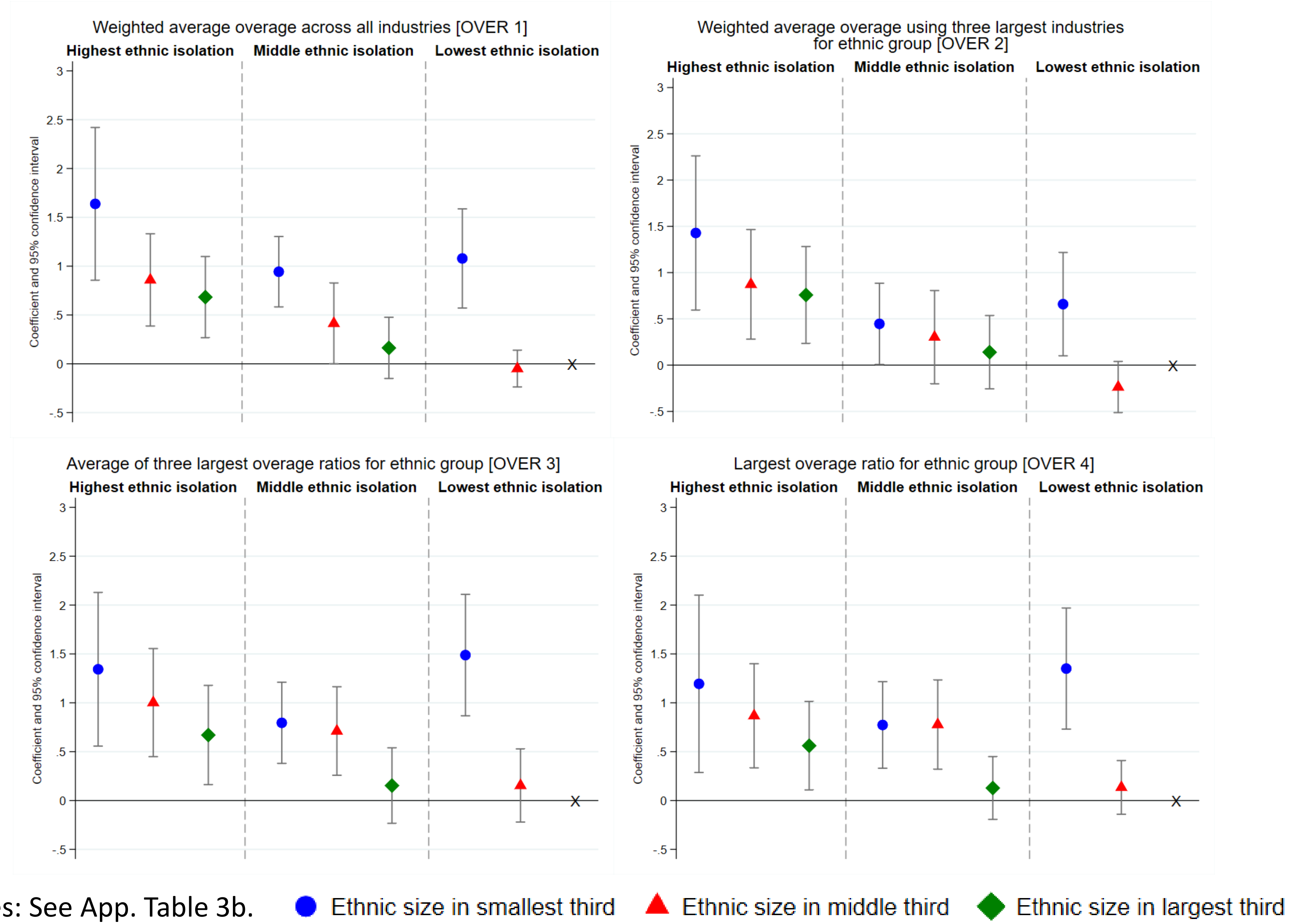
Notes: See Table 7.

App. Figure 1a: Non-parametric estimations without controls



Notes: See App. Table 3a. ● Ethnic size in smallest third ▲ Ethnic size in middle third ◆ Ethnic size in largest third

App. Figure 1b: Non-parametric estimations with controls



Notes: See App. Table 3b.

Online Appendix: Theory

The theory in this paper consists of two fundamental building blocks. First, social interactions and production are complementary. Second, different social relationships are not close substitutes for one another. The former is analyzed in the main text, and this appendix begins with additional discussion. We then consider pricing equilibrium and social networks with endogenous matching. The numbering of assumptions and propositions continues from the main text.

1 Discussion of Baseline Model

1.1 Quality and Convex Productivity

In addition to the quantity of social interactions with other self-employed entrepreneurs, the quality of these interactions could also matter for productivity. Let individual productivity for self-employed entrepreneurs in industry 1 increase both in the quantity and average productivity of other entrepreneurs in the sector of the same group. Write this as

$$\theta = \phi + \delta X_l \overline{\theta}, \tag{1}$$

where $\phi > 0$ is a productivity term, $0 < \delta < 1$ is a social multiplier, X_l is the fraction of entrepreneurs in group l, and $\overline{\theta}$ is the average productivity of these entrepreneurs. Solving for equilibrium productivity by setting θ equal to $\overline{\theta}$, individual productivity in group l is a function:

$$\theta\left(X_l\right) = \frac{\phi}{1-\delta X_l}. \tag{2}$$

Under these conditions, productivity is convex in the degree of specialization when taking both the quantity and the quality of interaction into account.[1] With this result in mind, we make the following assumption:

Assumption 1B *Productivity of self-employed entrepreneurs in industry* 1 *is convex in specialization:* $\theta'' > 0$.

Assumption 1B allows a full characterization of the efficient solution without having to resort to explicit functional form. We discuss further below. Convex productivity gives the following result:

[1] This specification highlights the differences from a standard interaction model. The standard model is generally specified so that individual productivity is a function of a group-specific term ϕ and the discounted mean of the group, $\delta\overline{\theta}$. Solving $\theta = \phi + \delta\overline{\theta}$, interaction exacerbates the difference in ϕ across groups, $\theta = \frac{\phi}{1-\delta} > \phi$, but the degree of specialization X_l has no effect on productivity.

Lemma *If productivity is convex, both groups never work in both industries.*

Proof: Assume by contradiction that an efficient distribution (X_A, X_B) exists where $0 < X_l < 1$ for $l = \{A, B\}$. Consider a marginal change ϵ in the ethnic composition of self-employed entrepreneurs in industry 1 while holding fixed the overall number of said entrepreneurs M (and therefore also the outputs of both industries). Taking the derivative of Q_1 with respect to ϵ, and evaluating it at $\epsilon = 0$:

$$\frac{\partial Q_1}{\partial \epsilon}\left(X_A + \frac{\epsilon}{N_A}, X_B - \frac{\epsilon}{N_B}\right) = \theta(X_A) + X_A\theta'(X_A) - \theta(X_B) - X_B\theta'(X_B) \quad (3)$$

Since (X_A, X_B) is efficient, and since X_l is interior, this derivative has to be zero.[2] But with convex productivity the derivative is zero only at $X_A = X_B$, which is the global minimum. This contradicts efficiency. ■

The efficient economy aims for maximum ethnic homogeneity in self-employed entrepreneurship in industry 1. Ruling out that both groups work in both sectors implies that only the specialized distributions along the two curves depicted in Figure 1 of the main text could possibly coincide with the transformation frontier. The shape of the entire transformation frontier can therefore be deduced by tracing out the maximum of the two curves in that figure.

Proposition 2 *If productivity is convex, there is a cutoff value v^* such that for $v < v^*$, the minority group specializes as self-employed entrepreneurs in industry 1, whereas for $v > v^*$, the majority specializes.*

Proof: Direct from Proposition 1 and Lemma proofs with convexity. ■

The right panel of Figure 1 of the main text also shows how the degree of specialization varies with the size of industry 1, as governed by v, and the cutoff value v^* for majority group specialization. The greater the value of v, the greater is the demand for industry 1 and the more people work in it. As industry 1 increases in size, the interaction externality generates a characteristic discrete jump from one type of equilibrium to another. At the point v^*, where many from group B have also joined self-employed entrepreneurship in industry 1, the economy abruptly moves from minority specialization to majority specialization.

[2] If the derivative is nonzero, then the output of industry 1 could increase while keeping the output of industry 0 constant. By subsequently increasing the number of workers in industry 0 marginally, a Pareto improvement is feasible, thus contradicting efficiency.

1.2 The Case of Non-Convex Productivity

To see that convexity is needed for the Lemma on ethnic homogeneity to hold, consider a non-convex production function where a threshold fraction must work as self-employed entrepreneurs in industry 1 for interaction to have value: $\theta > 0$ if $X_l \geq b$ and zero otherwise. This specification violates the assumption that productivity is strictly increasing in the degree of specialization. Then, if the demand for industry 1 output is so great that a single group cannot satisfy it entirely, $v > V(0,1)$, and if in addition $V(b,b) < v < V(b,1)$, efficiency requires that both ethnic groups work in both industries, contradicting the Lemma.

To see why, consider what would happen if one of the groups specialized completely. In this case the non-specialized group's degree of specialization would be positive but below b, causing the self-employed industry 1 entrepreneurs in that group to have zero productivity. If, however, the industrial distribution was unspecialized instead, with $X_A = X_B$, then self-employed industry 1 entrepreneurs in both groups would be as productive as those in the most productive group were under the alternative. Clearly this would be Pareto superior, contradicting the Lemma. This special case shows how the Lemma fails for non-convex productivity, and how in this case the qualitative features of specialization will depend on specific functional form assumptions. Recall however that the results for both $v \leq V(1,0)$ and $v = V(0,1)$ are more general and apply both for convex and non-convex productivity. This condition is less important for the remaining model discussion.

2 The Price Equilibrium

The model in the main text characterizes the efficient outcome. The focus now turns to the competitive outcome. An equilibrium analysis will yield two insights into how social interaction affects distribution over industries. First, it shows how stratifying forces act to make groups more and more different, and second, how group earnings are positively related to the degree of specialization.

To see how social interaction works as a stratifying force, begin by introducing time into the analysis, with $t = 0, 1, ..., \infty$. Dynamics are built into the model by making the interaction effect work with a lag. Denote by X_l^t the degree of specialization in period t for group l, and let self-employed individual entrepreneurial productivity in industry 1 in period t be a function $\theta\left(X_l^{t-1}\right)$. This one-period lag specification for the interaction effect could easily be generalized to a distributed lag. Interaction now effectively works as a form of social capital, with the group's self-employment activities in the previous period benefiting individual productivity today. Let p_1^t and p_0^t be the prices of industry 1 output and industry 0 output respectively. Entrepreneurial earnings in industry 1 are $y_{1,l}^t = p_1^t \theta\left(X_l^{t-1}\right)$ and worker earnings in industry 0 are $y_{0,l}^t = p_0^t$. Competitive industrial choice is straightforward to derive in this setting; defining the relative price of industry 0 output to industry 1 output as $p^t = \frac{p_0^t}{p_1^t}$, an individual in group l joins industry 1 as a self-employed entrepreneur if

$$\theta\left(X_l^{t-1}\right) \geq p^t \tag{4}$$

and favors being a worker in industry 0 if $\theta\left(X_l^{t-1}\right) \leq p^t$. Since individuals have identical skills, aggregate labor supply for group l is discontinuous, with:

$$X_l^t = \begin{cases} 1 & \text{if } \theta\left(X_l^{t-1}\right) > p^t \\ [0,1] & \text{if } \theta\left(X_l^{t-1}\right) = p^t \\ 0 & \text{if } \theta\left(X_l^{t-1}\right) < p^t. \end{cases} \tag{5}$$

Avoid for now the knife-edge *unspecialized* case where $X_A^{t-1} = X_B^{t-1}$. Since there is a single price of labor, p^t, at least one of the two groups A and B must then be in a corner:

$$\left(X_A^t, X_B^t\right) = \begin{cases} \left(X_A^t = 1, 0 < X_B^t\right) \text{ or } \left(X_A^t \leq 1, X_B^t = 0\right) & \text{if } X_A^{t-1} > X_B^{t-1} \\ \left(0 < X_A^t, X_B^t = 1\right) \text{ or } \left(X_A^t = 0, X_B^t \leq 1\right) & \text{if } X_A^{t-1} < X_B^{t-1} \end{cases} \tag{6}$$

In equilibrium, supply must satisfy (6) and production must meet demand so that markets clear. Because of perfect complementarity, meeting demand reduces to satisfying $v = V\left(X_A^t, X_B^t\right)$. The resulting equilibrium distribution is unique. To see why, take the case when group l is more specialized than group l' in the previous period, with $X_l^{t-1} > X_{l'}^{t-1}$. Given that at least one of the two groups must be in a corner according to (6), the equilibrium distribution must either be of the type $\left(X_l^t, 0\right)$ or of the type $\left(1, X_{l'}^t\right)$. Since the function V is strictly increasing in both arguments, it follows that $V\left(1, X_{l'}^t\right) > V\left(X_l^t, 0\right)$. Only one distribution can consequently make V equal to v.

The equilibrium distribution is therefore uniquely determined by the distribution in the previous period. Continuing to avoid the knife-edge unspecialized case, define a function ϕ that maps every previous distribution into a new distribution:

$$\left(X_A^t, X_B^t\right) = \phi\left(X_A^{t-1}, X_B^{t-1}\right) \tag{7}$$

Next, proceed to characterize stationary equilibrium distributions. Like other equilibrium distributions, stationary distributions must satisfy (6) and must meet demand. Following the same argument as above, based on V being strictly increasing in both arguments, it follows that there is a stationary equilibrium where each of the two groups specializes. Denote the stationary distribution as $\left(X_A^A, X_B^A\right)$ when the minority specializes, and the stationary distribution as $\left(X_A^B, X_B^B\right)$ when the majority specializes.

Finally, returning for a moment to the unspecialized knife-edge case where $X_A^{t-1} = X_B^{t-1}$, this type of initial condition is of measure zero and therefore not elaborated on. Note only that since V is strictly increasing in both arguments, there can only be one such stationary unspecialized equilibrium distribution. Denote that equilibrium distribution as $\left(X_A^U, X_B^U\right)$. In the unspecialized case, although there is only one stationary equilibrium, the uniqueness of equilibria no longer applies. To summarize, there are consequently three stationary equilibrium distributions: two specialized, $\left(X_A^A, X_B^A\right)$ and $\left(X_A^B, X_B^B\right)$, and one unspecialized, $\left(X_A^U, X_B^U\right)$. Figure A1 shows the two specialized equilibria, as well as the knife-edge equilibrium, when v is less than $V(1,0)$.

2.1 Industrial Stratification

Our next analysis shows that the dynamic system in (7) converges to a stationary specialized equilibrium, so long as the interaction externality is not too strong. This analysis only examines unspecialized initial conditions, which establishes convergence on measure one. Consider what happens to the aggregate production of industry 1 when one (infinitesimal) person in group l becomes a self-employed entrepreneur in that industry. First, aggregate production increases by an amount equal to the individual productivity of that person, $\theta(X_l)$. In addition, all other self-employed entrepreneurs in industry 1 from group l benefit from the interaction externality when socializing with this new entrepreneur. Individual productivity therefore increases by $\frac{1}{N_l}\theta'(X_l)$ for all $X_l N_l$ self-employed industry 1 entrepreneurs in group l. Consequently, the internalized effect on aggregate production of one person joining the self-employed entrepreneurial sector of industry 1 is $\theta(X_l)$, and the external effect is $X_l\theta'(X_l)$. Assume that the external effect is smaller than the internal effect.[3]

Assumption 2 *The internal effect dominates:* $\theta'(X_l)X_l < \theta(X_l)$.

This condition is satisfied if productivity is concave in X_l, but it also holds for some convexity as long as $\theta(0) > 0$. To see why the assumption is needed for the system to be stable, consider the extreme case when group A has no mass at all, with $N_A = 0$. Since the derivative of V with respect to X_A^t is zero in this case, group A can be ignored altogether in the general equilibrium analysis. There is then a single stationary level of specialization for group B; denote this value as X_B^*.

Consider a perturbation in period t so that the majority starts out with too many entrepreneurs in industry 1, $X_B^t > X_B^*$, shown in Figure A2. Such a deviation boosts the interaction effect in period $t+1$ relative to the stationary equilibrium, $\theta(X_B^t) > \theta(X_B^*)$. With perfect complementarity, the outputs of both industry 0 and industry 1 must therefore increase relative to their stationary equivalents. Increasing the output of industry 0 requires an increase in the number of workers in that industry, and consequently, a decrease in the number of self-employed entrepreneurs in industry 1 to

[3] We thank Rachel Soloveichik for this interpretation of Assumption 2.

below the stationary value X_B^*. With fewer of these entrepreneurs in period $t+1$ than the stationary number, the tables turn in period $t+2$, so that the interaction effect now is reduced to below that in the stationary equilibrium. Reducing the production of industry 0 and industry 1 in period $t+2$ in response, the number of industry 0 workers in period $t+2$ has to decrease and the number of self-employed industry 1 entrepreneurs has to increase relative to the stationary equilibrium. These reversals repeat every period in cobweb-style dynamics.[4]

The question of whether the system is stable reduces to whether the number of self-employed entrepreneurs in industry 1 in period $t+2$ is less than the number of such entrepreneurs in period t, so that the degree of specialization in group B gets closer and closer to the stationary value X_B^* over time. Using the derived direction of the change in industry 1 production, $Q_1^{t+1} > Q_1^{t+2}$, this latter inequality can be equivalently expressed, after multiplying and dividing the left-hand side by X_B^t and dividing both sides by $X_B^{t+1} N_B$, as:

$$X_B^t \frac{\theta\left(X_B^t\right)}{X_B^t} > X_B^{t+2} \frac{\theta\left(X_B^{t+1}\right)}{X_B^{t+1}} \tag{8}$$

Given that productivity is not too convex, as stipulated by Assumption 2, it follows that $\frac{\theta(X_l)}{X_l}$ is strictly decreasing in X_l. Since $X_B^t > X_B^{t+1}$, equation (8) then establishes that $X_B^t > X_B^{t+2}$. This proves convergence and the stability of group B's degree of specialization around X_B^*.

Having established stability in the case of $N_A = 0$, the same example also serves to show how the stratifying force comes into play. Let group B be in its stable state, with $X_B^t = X_B^*$, and perturb the minority's industry distribution so that $X_A^t > X_B^*$. Since group B is so much greater in size than group A, the former is unaffected by the perturbation and the price continues to be locked in at $p^{t+1} = \theta\left(X_B^*\right)$. The interaction effect in period $t+1$, generated by the perturbation in period t, then results in everyone in group A becoming more productive as self-employed entrepreneurs in industry 1 than as workers in industry 0, with $\theta\left(X_A^t\right) > p^{t+1}$. Group A's degree of specialization consequently jumps from X_A^t to $X_A^{t+1} = 1$, and the distribution stays in this stratified state forever. This stratification result is extended later for the general case of any population size of the two groups, and it follows that for $l \in \{A, B\}$ and $l' \in \{A, B\}$:

Proposition 3 *Initial differences result in long-run specialization: If group l is more specialized than group l' initially, $X_l^0 > X_{l'}^0$, then group l specializes in the long run and the limiting distribution is $\left(X_A^l, X_B^l\right)$.*

[4]The flip-flopping character of the equilibrium distribution is a result of the one-period lag specification for the interaction effect. The distribution would change more gradually with a more general specification allowing for distributed lags.

Proof: Consider the equilibrium sequence of industry distributions:

$$\left(\left(X_A^1, X_B^1\right), \left(X_A^2, X_B^2\right), ...\right) \tag{9}$$

If one group l is more specialized than the other group l' initially, $X_l^0 > X_{l'}^0$, supply in (5) requires that the equilibrium sequence begins in one of the following three ways:

$$\left(\left(X_l^1, X_{l'}^1\right), \left(X_l^2, X_{l'}^2\right), ...\right) = \begin{cases} \left((<1, 0), ...\right) \\ \left((1, \geq 0), (1, \geq 0), ...\right) \\ \left((1, \geq 0), (<1, 0), ...\right). \end{cases} \tag{10}$$

The proof proceeds by establishing that the sequence converges to $\left(X_A^l, X_B^l\right)$ in each of these three cases. Define the variable $\lambda(X_l) \equiv \frac{\theta(X_l)}{X_l}$ for $X_l > 0$. From Assumption 2 it follows that $\lambda'(X_l) < 0$. Proceed to establish convergence:

Case 1 $X_l^1 < 1$ *and* $X_{l'}^1 = 0$.

Show first that group l' stays out of entrepreneurship in industry 1 for good. By contradiction: if not, then there exists a time t where $X_{l'}^{t+1} = 0$ and $X_{l'}^{t+2} > 0$. Since supply must satisfy (6) it then follows that $X_l^{t+1} > 0$ and $X_l^{t+2} = 1$. The change in the output of industry 1 can then be written as:

$$Q_1^{t+2} - Q_1^{t+1} = N_l\left(\theta\left(X_l^{t+1}\right) - X_l^{t+1}\theta\left(X_l^t\right)\right) + X_{l'}^{t+2}N_{l'}\theta\left(X_{l'}^{t+1}\right). \tag{11}$$

This difference is strictly positive if the first term is positive. Clearly this is the case if $X_l^{t+1} \geq X_l^t$. If, instead, $X_l^{t+1} < X_l^t$, then again focusing on the first term:

$$\begin{aligned} \theta\left(X_l^{t+1}\right) - X_l^{t+1}\theta\left(X_l^t\right) &= \lambda\left(X_l^{t+1}\right)X_l^{t+1} - X_l^{t+1}\lambda\left(X_l^t\right)X_l^t \\ &= X_l^{t+1}\left(\lambda\left(X_l^{t+1}\right) - \lambda\left(X_l^t\right)X_l^t\right) > 0. \end{aligned} \tag{12}$$

This establishes that $Q_1^{t+2} > Q_1^{t+1}$. Since the output production of both industries must move in the same direction to clear the market, because of perfect complementarity, it follows that the output of industry 0 also increases from $t+1$ to $t+2$. This in turn requires that the number of workers in industry 0 increases, or equivalently, that the number of self-employed entrepreneurs in industry 1 decreases:

$$X_l^{t+2}N_l + X_{l'}^{t+2}N_{l'} < X_l^{t+1}N_l + X_{l'}^{t+1}N_{l'}. \tag{13}$$

Since $X_l^{t+2} = 1$ and $X_{l'}^{t+1} = 0$, this inequality can be simplified as $N_l + X_{l'}^{t+2}N_{l'} < X_l^{t+1}N_l$. This inequality is a contradiction and establishes that group l' stays out of self-employed entrepreneurship in industry 1 for good. The stationary equilibrium must consequently be of the form $\left(X_l^l, 0\right)$.

Assume first that $X_l^t > X^*$, in which case it is easy to show that $Q_1^{t+1} > Q_1^l > Q_1^{t+2}$ as well as $X_l^{t+1} < X_l^l < X_l^{t+2}$. Since $Q_1^{t+1} > Q_1^{t+2}$ it follows that:

$$\begin{aligned} X_l^{t+1} N_A \theta\left(X_l^t\right) &> X_l^{t+2} N_A \theta\left(X_l^{t+1}\right) \\ X_l^{t+1} \lambda\left(X_l^t\right) X_l^t &> X_l^{t+2} \lambda\left(X_l^{t+1}\right) X_l^{t+1} \\ X_l^t \lambda\left(X_l^t\right) &> X_l^{t+2} \lambda\left(X_l^{t+1}\right). \end{aligned} \tag{14}$$

The last line implies that $X_l^t > X_l^{t+2}$. The exact same argument, but with reverse inequalities, can be made for $X_l^t < X_l^l$. Therefore, having established that $X_l^t > X_l^{t+2} > X_l^l$ when $X_l^t > X_l^l$, and vice versa when $X_l^t < X_l^l$, it has been shown that X_l^t approaches the stationary equilibrium value X_l^l over time. This establishes convergence in Case 1.

Case 2 $X_l^1 = 1,\ X_{l'}^1 \geq 0,\ X_l^2 = 1$ *and* $X_{l'}^2 \geq 0.$

Show first that in this case, group l stays specialized for good. By contradiction: if not, then there exists a time t when $X_l^t = 1$, $X_l^{t+1} = 1$ and $X_l^{t+2} < 1$. Since supply must satisfy (6), it follows that $X_{l'}^{t+2} = 0$. The change in the output of industry 1 can be written as

$$Q_1^{t+2} - Q_1^{t+1} = N_l\left(X_l^{t+2}\theta(1) - \theta(1)\right) - X_{l'}^{t+1} N_{l'} \theta\left(X_l^t\right) < 0. \tag{15}$$

Since the supply of output of both industries must move in the same direction to clear the market, it follows that the output of industry 0 also decreases, which requires that the number of self-employed entrepreneurs in industry 1 increases:

$$X_l^{t+2} N_l + X_{l'}^{t+2} N_{l'} > X_l^{t+1} N_l + X_{l'}^{t+1} N_{l'}. \tag{16}$$

Since $X_{l'}^{t+2} = 0$ and $X_l^{t+1} = 1$, this inequality can be rewritten as $X_l^{t+2} N_l > N_l + X_{l'}^{t+1} N_{l'}$, which is a contradiction. This establishes that group l stays specialized in industry 1 for good. The stationary equilibrium must consequently be of the form $\left(1, X_{l'}^l\right)$. By the same argument as in Case 1, the sequence can be shown to approach the stationary equilibrium value $X_{l'}^l$ over time, both if $X_{l'}^t > X_{l'}^l$ and if $X_{l'}^t < X_{l'}^l$. This establishes convergence in Case 2.

Case 3 $X_l^1 = 1$ *and* $X_{l'}^1 \geq 0$ *and* $X_l^2 < 1$ *and* $X_{l'}^2 = 0.$

By the same argument in Case 1, it follows that group l' stays out of entrepreneurship in industry 1 permanently. Repeating the arguments in Case 1, convergence can then be established also in Case 3.

Consequently, in all three cases there is convergence. ■

This also implies that the stationary unspecialized equilibrium $\left(X_A^U, X_B^U\right)$ is unstable. If the minority group is slightly more specialized initially, then the economy converges to minority specialization $\left(X_A^A, X_B^A\right)$, and if the opposite is true, then the economy converges to majority specialization $\left(X_A^B, X_B^B\right)$. Over time, social segregation amplifies initial group differences.

2.2 Initial Conditions and Multiple Groups

Depending on the initial conditions, as is clear from Proposition 3, either of the two groups A and B can specialize as self-employed entrepreneurs in industry 1. Social interaction amplifies initial differences, but it does not explain why they are there to begin with. The difference in group size has some implications for what initial conditions to expect, however.

Consider an economy with more than two groups. As before, the group with more self-employed entrepreneurs in industry 1 initially will specialize in the long run. If the initial industrial distribution is subject to randomness, one of the smaller groups is likely to be the most specialized initially. To see why, let the initial distribution be generated by random draws, where each person becomes a self-employed entrepreneur in industry 1 with probability ρ.[5] This probability structure results in the same expected initial degree of specialization for all groups, but since the population size varies across groups, the variance in the degree of specialization also varies. The smallest groups have the largest variance, and therefore, the smallest groups are most likely to exhibit the lowest and also the greatest initial degrees of specialization. Consequently, with the smallest groups the most likely to specialize initially, as interaction amplifies initial differences over time, the smallest groups are also the most likely to specialize in the long run.

2.3 Assimilation

Our model does not feature assimilation of immigrants and their offspring and thus yields permanent social and industrial segregation. In our framework, assimilation would reduce the social isolation of an ethnic group (or some members of it) to the majority group. Our framework then predicts the industry choices of the assimilated individuals to look like those of the majority, especially if another ethnic group shows strong social isolation.

[5] These draws can be partially correlated within groups with the assumption that the correlation is the same for every group.

2.4 Heterogeneity and Earnings

Social complementarities also have implications for earnings. To examine how interaction effects would show up in earnings data, it is necessary to move away from the framework of identical skills. Returning to a static environment, endow each person i with entrepreneurial skills relevant to self-employment in industry 1, $s_1(i)$, and with another set of skills necessary for industry 0, $s_0(i)$. Self-employed entrepreneurial earnings in industry 1 are now a function of both interactions and skills. Denote the earnings of individual i in group l when she is a self- employed entrepreneur in industry 1 as $y_1(X_l, i) = p_1 \theta(X_l) s_1(i)$, and when she is a member of industry 0 as $y_0(i) = p_0 s_0(i)$. Defining the ratios $s \equiv \frac{s_1}{s_0}$, $p \equiv \frac{p_0}{p_1}$, and $q \equiv p\frac{y_1}{y_0}$, the earnings-maximizing industry choice of individual i is to consider becoming a self-employed entrepreneur in industry 1 if:

$$q(X_l, i) \geq p \tag{17}$$

and to consider working in industry 0 if $q(X_l, i) \leq p$. Here the term $q(X_l, i) = \theta(X_l) s(i)$ summarizes the individual's comparative advantage in self-employed entrepreneurship in industry 1, at parity prices, as a function of social interaction and skills.

When individuals have different skills, the character of the price equilibrium depends crucially on the marginal self-employed entrepreneur and how her comparative advantage changes as more and more untalented people also become entrepreneurs in industry 1. If the benefits of interaction are weak and the marginal entrepreneur "deteriorates" as more intrinsically untalented people enter the industry, then the economy reduces to a standard Roy model, or sorting model, with a unique unspecialized equilibrium. Only if the interaction effect is strong enough to overcome skill heterogeneity can interaction change the character of the equilibrium.

Without loss of generality, order individuals from the greatest to the smallest comparative advantage in industry 1-style entrepreneurship, so that the skill ratio is decreasing in i, $s'(i) \leq 0$. The marginal entrepreneur is then the individual indexed by $i = X_l$, and her comparative advantage is $q(X_l, X_l)$. To prevent the economy from reducing to a sorting model, assume that the interaction effect trumps heterogeneity:

Assumption 3 *Interaction dominates at the margin:* $\frac{d}{dX_l} q(X_l, X_l) > 0$.

This assumption implies that the solid line in Figure A3 is upward sloping. The equilibrium distribution (X_A, X_B) must be competitively supplied and enough output must be produced by both industries to meet demand. Using a similar line of reasoning as in the previous section, based on V being strictly increasing in both arguments, it follows from Assumption 3 that there are three equilibria: one unstratified, denoted $\left(X_A^U, X_B^U\right)$; one where the minority group A specializes, denoted $\left(X_A^A, X_B^A\right)$; and one where the majority group B specializes, denoted $\left(X_A^B, X_B^B\right)$.[6]

In the equilibrium where minority A specializes as self-employed entrepreneurs in industry 1, the mean earnings of members of group A are higher than the mean earnings of members of group B, and vice versa in the equilibrium where group B specializes. To see why, let $y = \max(y_0, y_1)$ be actual individual earnings, and denote mean group earnings as $\mu = \int_0^1 y di$.

Proposition 4 *Earnings covary with self-employed entrepreneurship in industry 1:* $\mu(X_l) > \mu(X_{l'})$ *if* $X_l > X_{l'}$.

Proof: Since people sort into industries, mean earnings can be rewritten as

$$\mu(X_l) = \int_0^1 y_0(i)\, di + \int_0^{X_l} \left(y_1(X_l, i) - y_0(i)\right) di \tag{18}$$

Rearranging, the difference in mean earnings between the two groups is:

$$\mu(X_l) - \mu(X_{l'}) = \int_0^{X_{l'}} \left(y_1(X_l, i) - y_1(X_{l'}, i)\right) di + \int_{X_{l'}}^{X_l} \left(y_1(X_l, i) - y_0(i)\right) di \tag{19}$$

where both parts of the expression are positive. The first part is strictly positive due to the interaction effect, $\frac{\partial y_1(X_l,i)}{\partial X_l} > 0$, and the second part is positive because of sorting, $y_1(X_l, i) \geq y_0(i)$ for all $i \leq X_l$. ■

This unequivocal effect on mean earnings at the group level does not carry through to the industry level. Depending on the joint distribution of skills, mean earnings in either industry can increase or decrease as interaction increases self-employed entrepreneurial productivity in industry 1 and shifts people of different ability between industries. The effect of interaction on industry earnings is similar to the effect of changing skill prices, which cannot be signed for a general skill distribution (Heckman and Honore, 1990).

The difference in mean earnings, normalized in units of industry 0 output, is shown in Figure A4 for the equilibrium with minority specialization. The exact derivation is included below. The relative price of industry 0 to industry 1 outputs is always such that the marginal entrepreneur is indifferent between industries. Keeping track of whether the marginal entrepreneur is in group A or in group B depending on the industrial distribution, the equilibrium price can be expressed as:

$$p = \begin{cases} q(X_l, X_l) & \text{if } X_l > X_{l'} \text{ and } X_{l'} = 0, \text{ or } X_l < X_{l'} \text{ and } X_l > 0 \\ q(X_{l'}, X_{l'}) & \text{if } X_l > X_{l'} \text{ and } X_{l'} > 0, \text{ or } X_l < X_{l'} \text{ and } X_l = 0 \end{cases} \tag{20}$$

[6] Note that Assumptions 2 and 3, when combined, put both an upper and a lower bound on the interaction effect: $-\frac{d \ln s}{dX_l} < \frac{d \ln \theta}{dX_l} < \frac{1}{X_l}$.

When increasing the number of self-employed entrepreneurs in industry 1 in equilibrium with minority specialization, the relative price of industry 0 output to industry 1 output increases continuously as the marginal entrepreneur in group A becomes more and more productive. This increase in price continues until all As are self-employed entrepreneurs in industry 1. To expand industry 1's self-employed entrepreneurial sector further from the point where everyone in group A are entrepreneurs, the price has to drop discretely from $p = q(1,1)$ to $q(0,0)$, to lure the unproductive Bs into the sector as well. The earnings differential between groups A and B moves accordingly, as shown in Figure A4, increasing continuously until all As are self-employed entrepreneurs in industry 1, at which point earnings jump in response to the discontinuous drop in the relative price.

Derivation of Earnings Differential in Figure A4: Mean earnings denominated in terms of industry 0 outputs are:

$$\frac{\mu(X_l)}{p_0} = \int_0^{X_l} p^{-1}\theta(X_l)\, s_1(i)\, di + \int_{X_l}^{1} s_0(i)\, di. \tag{21}$$

Replace the relative price of industry 0 output to industry 1 output, $p = \frac{p_0}{p_1}$, with the comparative advantage of the marginal entrepreneur, q, since these two are equal in equilibrium. Denote the earnings differential as $\Delta(X_l, X_{l'}) \equiv \frac{\mu(X_l) - \mu(X_{l'})}{p_0}$. It can be expressed as:

$$\Delta(X_l, X_{l'}) = \int_0^{X_{l'}} q^{-1}\left(\theta(X_l) - \theta(X_{l'})\right) s_1(i)\, di + \int_{X_{l'}}^{X_l} \left[q^{-1}\theta(X_l)\, s_1(i) - s_0(i)\right] di. \tag{22}$$

For $X_l < 1$ and $X_{l'} = 0$, where $q = q(X_l, X_l)$, and $q(X_l, X_l) = \theta(X_l)\, s(X_l)$, differentiating with respect to X_l gives

$$\frac{\partial \Delta(X_l, 0)}{\partial X_l} = -s'(X_l)\, s(X_l)^{-2} \int_0^{X_l} s_1(i)\, di > 0. \tag{23}$$

For $X_l = 1$ and $X_{l'} = 0$, the drop in price from $q(1,1)$ to $q(0,0)$ results in a jump in the mean earnings differential equal to

$$\Delta(1,0)\big|_{p=q(0,0)} - \Delta(1,0)\big|_{p=q(1,1)} = \left(q(0,0)^{-1} - q(1,1)^{-1}\right)\theta(1) \int_0^1 s_1(i)\, di > 0. \tag{24}$$

For $x = 1$ and $X_{l'} > 0$, where $q = q(X_{l'}, X_{l'})$, differentiating with respect to $X_{l'}$ gives

$$\frac{\partial \Delta(1, X_{l'})}{\partial X_{l'}} = -\frac{dq}{dX_{l'}} q^{-2}\theta(1) \int_0^1 s_1(i)\, di + s'(X_{l'})\, s(X_{l'})^{-2} \int_0^{X_{l'}} s_1(i)\, di - 2 s_0(X_{l'}) < 0. \tag{25}$$

■

3 Relationships in a Social Network

Since interactions have been restricted to be random, the analysis has so far abstracted from changes in the social structure that could arise in response to the productive value of interaction. The most interesting question is whether the majority will split up into smaller social groups, formed around choice of industry, to capitalize on interaction. If such splinter groups could form *costlessly*, then social interaction would no longer be able to generate industrial stratification along ethnic lines.

By developing a utility-based theory of interaction, explicitly stating social preferences and characterizing the optimal social structure, this section shows that splinter groups will not arise so long as preferences are sufficiently diverse, and so long as different social relationships are not close substitutes for one another. Under these two premises it is costly to confine social interactions to within a small group since the quality of social matches deteriorates with decreasing group size.

The theory developed in this section is constructed around a standard marriage market as in Becker (1973). In addition to spousal matching, people are also related by birth, which yields a larger social structure where individuals are interrelated not just pairwise but in a social network. Since the social network is derived as the outcome of matching, the problem analyzed here is different in nature from the problems most commonly analyzed in the social network literature, for example in Jackson and Wolinsky (1996), which focuses on strategic interaction between identical agents.

3.1 The Marriage Market

Take a very large finite population $i = 1, ..., N$, which is divided into mutually exclusive and exhaustive *families* by birth, with each family consisting of $d > 3$ individuals. Every person i independently draws a trait t_i, which could be for example beauty or intelligence, uniformly distributed between zero and one:

Assumption 4 *Individual traits t_i are independent draws.*

The independence of the draw signifies what can be thought of as maximal diversity: even within families people have different traits.

Based on realized traits, each person is assigned a spouse. To simplify, there are no gender restrictions and spouses can belong to the same family.[7] Traits are assumed to be complementary inputs in marriage. A marriage between i and j yields utility $u(t_i, t_j)$, where the function u is symmetric and strictly increasing with a positive cross-derivative:

[7] Removing gender restrictions maps this problem into a one-sided assortative matching problem. One-sided assortative matching is used in a different context in Kremer (1993).

Assumption 5 *Inputs are complementary:* $u(t_i, t_j) = u(t_j, t_i)$, $u_1 > 0$, $u_2 > 0$ *and* $u_{1,2} > 0$.

Since different relationships produce different utility, social relationships are not perfect substitutes and there is an optimal matching of spouses. Assume that utility is transferable, in which case the efficient spousal matching has to maximize aggregate utility. Labelling individuals according to rank, so that $t_1 < t_2 < ...$,[8] it follows that the efficient matching is positively assortative: person one marries person two, person three marries person four, ..., and person $N-1$ marries person N. To see this, let the matching function v be symmetric and the cross-derivative positive. For traits $t_1 < t_2 < t_3 < t_4$, we show that the only efficient matching is (t_1, t_2) and (t_3, t_4). As in Becker (1973), we use a property of v when the cross-derivative is positive,

$$v(a, d) + v(c, b) < v(a, b) + v(c, d) \tag{26}$$

for $a < c$ and $b < d$. Take an arbitrary efficient matching (x_1, x_2) and (x_3, x_4), which is a permutation of the traits t_1, t_2, t_3 and t_4. Without loss of generality, relabel these traits pairwise so that $x_1 < x_2$ and $x_3 < x_4$. Also without loss of generality, relabel the pairs so that $x_1 < x_3$. This implies that $x_1 < x_3 < x_4$. Using the symmetry of v, the aggregate utility from the arbitrary efficient matching can be written as $v(x_1, x_2) + v(x_4, x_3)$. Since $x_1 < x_4$ it follows from (26) that $x_2 < x_3$, otherwise aggregate utility could be increased by interchanging x_2 and x_3, just as b and d were interchanged in (26). Consequently, with $x_1 < x_2 < x_3 < x_4$, the arbitrarily chosen efficient matching (x_1, x_2) and (x_3, x_4) is identical to the efficient matching (t_1, t_2) and (t_3, t_4).

3.2 Splinter Groups

Say that two people i and j are *related* if they are married and/or belong to the same family. Define a *splinter group* as a proper subset of the population where no one in the subset is related to anyone outside of that subset. Given an efficient assignment of spouses in a very large population where traits are independently distributed, it follows that:

Proposition 5 *The probability that splinter groups exist is zero.*

Proof: Define a d-regular multigraph with loops, where every vertex corresponds to a family, and every edge corresponds to a marriage. A splinter group is equivalent to an unconnected component of this graph. Assortative marriages on independent traits generate a random configuration of vertices. A random configuration is equivalent to

[8] Since having equal-valued traits, $t_i = t_j$, is of measure zero, this possibility is ignored.

a regular random multigraph, as defined in Janson et al. (2000). A regular random multigraph is asymptotically almost surely Hamilitonian for $d > 3$ (Janson et al. 2000). Connectivity follows from Hamiltonicity, which rules out the existence of unconnected components, and consequently, the existence of splinter groups. ■

A partial explanation for this result is that if person i marries person j, then because of the independence of traits, it is unlikely that anyone else in i's family marries into j's family as well. As the population grows larger, it becomes less and less likely that there is more than one marriage between the families of i and j. This "mismatch" prevents i and j, and their families, from socially isolating themselves from the larger population. The problem is more interesting than what this partial intuition conveys, however. The likelihood of more than one marriage between two particular families decreases as the population grows larger, but on the other hand, the number of families for whom this event could occur increases. If, for example, d had been equal to two, then these two effects would have balanced, so that small splinter groups would have formed even as the population approached infinity. This proof most likely also goes through for $d \geq 3$, since it really only needs connectivity and since connectivity is closely related to cubic graphs. The fourth edge is necessary in the case of multigraphs to ensure Hamiltonicity, but Hamiltonicity is stronger than connectivity.

In addition to the above proof, we can provide a more structured intuition for no splinter groups by using a branching tree to trace out relationships in the population. Let Σ be the set of all families. Define an arbitrary family in Σ as the singleton set $\sigma(0)$. Let $\sigma(1)$ be the set of families in $\Sigma/\sigma(0)$ with at least one family member married to someone in the original family $\sigma(0)$. Define $\sigma(2)$ as the set of families in $\Sigma/(\sigma(0) \cup \sigma(1))$ with at least one family member married to someone in $\sigma(1)$. Continuing by iteration to more and more distant relations, let $\sigma(r)$ be the set of families in $\Sigma/(\sigma(r-2) \cup \sigma(r-1))$ married to someone in $\sigma(r-1)$. The variable r denotes what is sometimes called the degree of separation between the initial family $\sigma(0)$ and the families in $\sigma(r)$. The degree of separation is a measure of the social distance between individuals; compare Milgram (1967). The collection of these sets, $\cup_{q=0}^{r} \sigma(q)$, constitutes a branching tree. The sets in this collection are mutually exclusive, but if there are splinter groups, the sets are not exhaustive even as $r \to \infty$. Denote by $s(r)$ the cardinality of the set $\sigma(r)$. Since each family in $\sigma(r)$ is composed of d family members, where at least one member in each family by definition is married into $\sigma(r-1)$, the expansion of the tree $\cup_{q=0}^{r} \sigma(q)$ is bounded by

$$s(r+1) \leq s(r)(d-1). \tag{27}$$

If equation (27) holds with equality, then as r increases $s(r)$ very soon encompasses the entire population. It turns out that the equation generally holds as an inequality, however. The reason for this slowdown is threefold. First, a person in $\sigma(r)$ could marry another person in $\sigma(r)$. Second, a family in $\sigma(r)$ could have more than one family member married to someone in $\sigma(r-1)$. Thirdly, several people in $\sigma(r)$ could marry into the same family. These three types of events combine to prevent each family in $\sigma(r)$ from contributing a full $d-1$ new families to $\sigma(r+1)$, and consequently cause (27) to hold as an inequality.

Applying the branching tree $\cup_{q=0}^{r} \sigma(q)$ to the efficient assortative matching, the branching tree is overwhelmingly likely to grow to encompass the entire population in the limit. Since the branching tree only expands to include people who are directly or indirectly related, this limit result is equivalent to Proposition 5 that there are no splinter groups. To see why the entire population is included in the limit, consider what would happen if it were not true, if the branching tree died out without having reached a positive fraction of the population. If this were the case, then $\sigma(r)$ would eventually have to grow arbitrarily small relative to the remainder set $\Sigma / (\sigma(r-2) \cup \sigma(r-1))$, and therefore the likelihood that someone in $\sigma(r)$ married someone else in $\sigma(r)$ rather than in the remainder set, or that several people in $\sigma(r-1)$ married into the same family in $\sigma(r)$ rather than in the remainder set, or that several people in $\sigma(r)$ married into the same family in the remainder set, must also grow arbitrarily small. But then equation (27) should hold as an equality, implying that $s(r+1) > s(r)$, which contradicts the premise that the branching tree died out without having reached the entire population. Consequently, everyone in the population is either directly or indirectly related, and there are no splinter groups.

3.3 Implications for Productivity

The social network developed here allows more individual choice than the random interaction model analyzed earlier, since here industry choice can be made contingent on every aspect of the social structure. The main results from the random interaction model continue to hold nevertheless. A large group cannot align social relationships so as to maximize productivity in a small industry where social interaction and productivity are complementary, without incurring the cost of deteriorating social matches that comes from breaking up into smaller groups. This follows from the result that no splinter groups arise under first-best matching on social traits. Since the social choice set of ethnic minority groups is restricted anyway, these groups can limit their social interactions to a single industry at no alternative cost. Ethnic minorities are therefore well suited for social interaction-intensive industries.

A social network with the same properties could also be derived from a meeting technology where spouses meet and marry at random. The social structure derived here can therefore equally well be thought of as arising in a rigid environment where people meet randomly, as arising from efficient matching. Since randomness is likely to also have a role in who marries whom, this adds additional strength to the result.

Breaking up into smaller groups not only carries a social utility cost but also carries the cost of having to bypass naturally occurring random matching.

3.4 Future Model Extensions

An interesting extension for future work is to include both general and specific skills in the same framework. In such a model of spillovers between sectors, it should be possible to derive stratification in overall entrepreneurial activity as well as industry stratification between different forms of self-employed entrepreneurship at the same time. This would correspond to the current situation in the United States, where groups like the Koreans are strongly clustered in a few business sectors, while at the same time being overrepresented as self-employed owners in almost all other business activities as well.

4 Additional Appendix References:

Becker, Gary. 1973. A Theory of Marriage: Part I. *Journal of Political Economy* 81: 813-846.

Heckman, James and Bo Honore. 1990. The empirical content of the Roy model. *Econometrica* 58: 1121-1149.

Jackson, Matthew, and Asher Wolinsky. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* 71: 44-74.

Janson, Svante, Tomasz Luczak, and Andrzej Rucinski. 2000. *Random Graphs.* New York: John Wiley.

Kremer, Michael. 1993. The O-ring theory of economic development. *Quarterly Journal of Economics* 108: 551-575.

Milgram, Stanley. 1967. The small world problem. *Psychology Today* 22: 61-67.

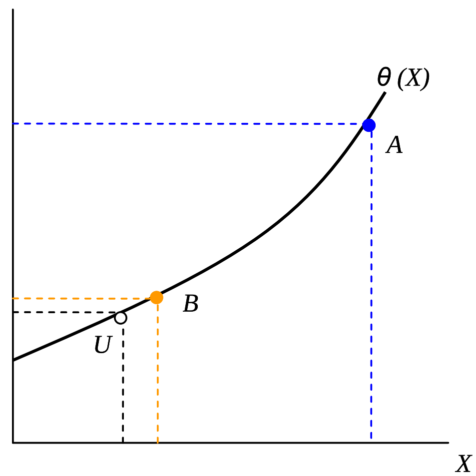


Figure A1. Individual productivity and the three stationary equilibria: one specialized equilibrium with minority specialization (A), one specialized equilibrium with majority specialization (B), and one unstratified equilibrium (U).

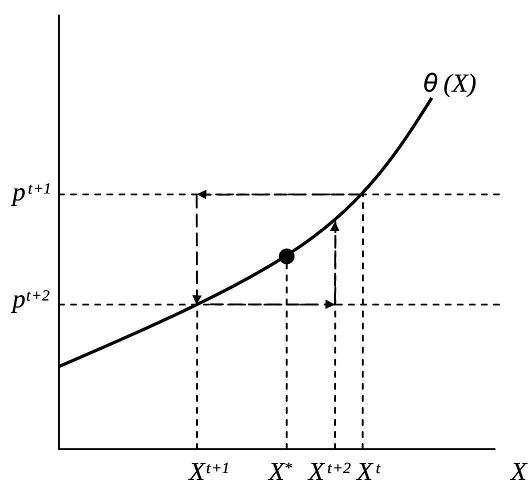


Figure A2. Stable dynamics when the internal effect dominates.

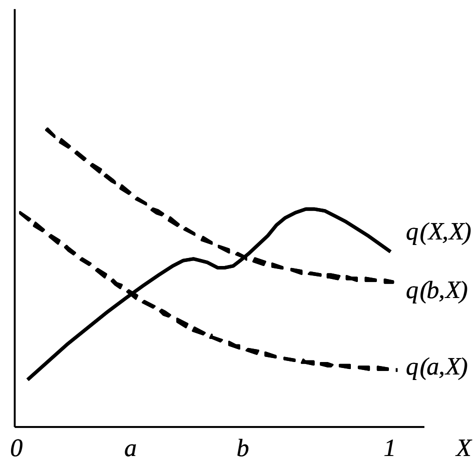


Figure A3. Sorting versus interaction effects in individual productivity. The dotted lines illustrate how the interaction effect raises productivity at all ability levels when specialization increases from *a* to *b*. The solid line shows the productivity of the marginal entrepreneur, for whom *i*=*X* at every level of *X*.

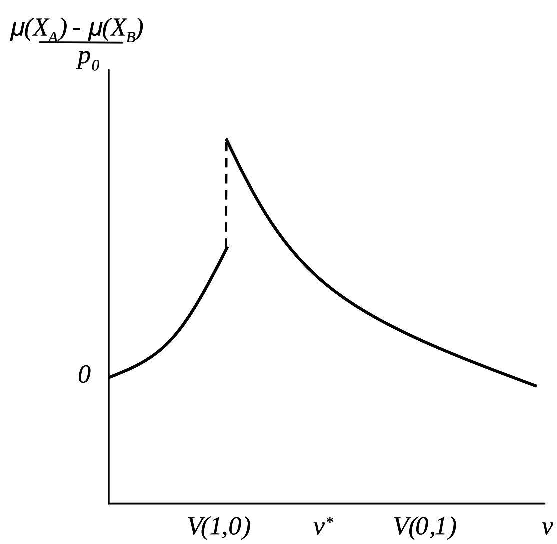


Figure A4. The difference in mean earnings between group *A* and group *B*, for different values of *v*, when minority group *A* specializes.

Online Appendix: Sociology Literature

Studies in sociology offer several theories and explanations for ethnic entrepreneurial specialization, including individuals' tie to their home country's culture, discrimination in the labor market, social cohesion theory, and cultural and/or religious traits and preferences. While it is beyond the paper's scope to offer a comprehensive review, this appendix introduces several connected theories as background for our study.

1 Connected Theories

A first set of work depicts the "sojourner status" of some migrant groups. These individuals are not planning to settle permanently in the host country to which they have temporarily moved and accordingly focus on the cultural heritage of their own ethnic group rather than assimilating into their host society (Siu, 1952). Sojourners may be inclined to seek portable occupations due to their expectation that they will return to their home country one day, and they have less incentive to invest in the business community outside of their ethnicity. Sojourners may be more reliant on their own group for business partnerships, hiring needs, and resources like capital investment and knowledge.

A connected "middleman minority" theory of ethnic employment is also based on individuals' tie with their home country. Middleman minorities are minority ethnic groups that tend to concentrate in intermediate occupations in their economies, facilitating transactions between one party and another. Examples of industries involving middleman minorities include trade and money lending (Blalock, 1967). Sojourner groups may end up in self-employment in middleman industries, which typically provide a portable livelihood requiring less fixed investment (Bonacich, 1973). Trust within the ethnic group may allow that group to compete successfully with native middleman businesses in the host country and encourage further specialization. Middleman minorities exhibit a preference for marriage within the ethnic group and established community institutions that reinforce the segregation of the ethnic group from those native to the host country (Bonacich, 1973). While important background, these theories have less direct application to our work that focuses on permanent migration and the industry specialization of ethnic entrepreneurs.

Whereas the initial theories depict a choice to become self-employed, others argue migrants can be forced into self-employment due to factors like discrimination that prevent them from accessing other labor opportunities (e.g., Wong, 1985) and the "blocked mobility hypothesis" (e.g., Light, 1972; Min, 1988a). Light and Gold (2000) write about the ethnic economy, the part of the labor market consisting of coethnic entrepreneurs and their coethnic employees. They summarize many advantages and disadvantages of ethnic economies for ethnic minorities, acknowledging the negative aspects of exploitation, discrimination, and inequalities while highlighting the positive benefits of increased opportunities for jobs, goods, and services that ethnic economies bring to an ethnic group in an economy.

The theory of “social cohesion” focuses on the direct and indirect social forces that act on members of a group to remain in a group. These forces include both individual group members’ attitudes and behaviors and the interactions between group members, which influence members’ attitudes and behaviors. Socially cohesive groups are often self-reinforcing; the greater the group cohesion, the more stable the group’s membership and the more group members are motivated to reinforce attitudes and behaviors that maintain group membership. Friedkin (2004) and Fonseca et al. (2019) provide reviews. Forces aiding social cohesion like within-group trust, information sharing, and cooperation could also facilitate the self-employment specialization of the ethnic group.

The related concept of “social capital” also provides potential explanations for immigrant self-employment. Social capital refers to resources gained through group connections that can be used for economic, social, or cultural purposes. Individuals’ can acquire social capital through participating in a group; social networks help generate reciprocity and trust. See Bourdieu (1986), Coleman (1988), and Putnam (1993) for varying definitions of social capital. As a member of an ethnic group develops more social capital, they may find it easier to start and maintain a business if it is in an industry others in their network are self-employed in, as they would have more access to information about business opportunities and more trust from potential business acquaintances and resources suppliers.

Concentrations of ethnic firms that occupy a particular urban space are known as ethnic enclaves. Studies of New York’s Chinatown, Miami’s Little Havana, and Los Angeles’ Koreatown emphasize the importance of social networks for obtaining start-up capital, business information, and access to the labor force. Similarly, an ethnic niche emerges when a group takes prominence in a sector of employment, where members find jobs for each other through network chains, and when entry-level openings are filled by kin and friends. Portes (1998) provides a review of both concepts.

Weak ties can be important for job referrals (Granovetter, 1973). Immigrant entrepreneurs, on the other hand, have particular use of strong ties with kin and co-ethnics. When ties are deeply embedded within their networks, they are more likely to engage in the receipt and transmission of business support and information (Waldinger et al., 1990). Light (1972), Min (1988a), and Bates (1997) consider access to credit within ethnic networks. Bonacich and Light (1988) study how Koreans in Los Angeles are brought together for ritualized occasions (e.g., church) and how they afterwards exchange information about business conditions and techniques; Kim (1987) considers similar patterns in New York. A group's ability to exploit opportunities is linked to their internal organizational capacity. Ethnic groups with densely connected networks can support aspiring business owners through friends and family and through ethnic institutions like religious organizations. Aldrich and Zimmer (1986) provides an overview.

Several researchers have also emphasized the importance of cultural and religious distinctions within an ethnic group to how individuals organized themselves to specialize in entrepreneurship. See, for example, the work of Morris (1956) on Indians in East Africa, Winder (1963) on Lebanese in West Africa, and Bonacich and Light (1991) and Min (1988a-c, 1990) on Koreans in the United States. Botticini and Eckstein (2005) delve into Jewish economic history to explain the occupational selection of Jewish people, theorizing that the occupational selection of Jews was due to the religious and educational reforms, which brought about widespread literacy and thereby a comparative advantage in starting businesses in skilled occupations as new urban centers developed. Sharma (2019) provides a recent UK depiction of cultural and social reasons for Asian immigrant specialization as shopkeeper entrepreneurs in the United Kingdom.

2 Additional Appendix References:

Aldrich, Howard, and Catherine Zimmer. 1986. Entrepreneurship through social networks. In D.L. Sexton and R.W. Smiler, eds. *The Art and Science of Entrepreneurship.* Cambridge, MA: Ballinger, 3-23.

Bates, Timothy. 1997. Financing small business creation: The case of Chinese and Korean immigrant entrepreneurs. *Journal of Business Venturing* 12(2): 109-124.

Blalock, Hubert. 1967. *Toward a Theory of Minority Group Relations.* New York: John Wiley.

Bonacich, Edna. 1973. A theory of middleman minorities. *American Sociological Review* 38: 583-594.

Botticini, Marestella, and Zvi Eckstein. 2005. Jewish occupational selection: Education, restrictions, or minorities? *Journal of Economic History* 65(4): 922-948.

Bourdieu, Pierre. 1986. The forms of capital. In J.G. Richardson, eds. *Handbook of Theory and Research for the Sociology of Education.* New York: Greenwood Press.

Coleman, James. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94: S95-S120.

Fonseca, Xavier, Stephan Lukosch, and Frances Brazier. 2019. Social cohesion revisited: A new definition and how to characterize it. *Innovation: The European Journal of Social Science Research* 32(2): 231-253.

Friedkin, Noah E. 2004. Social cohesion. *Annual Review of Sociology* 30: 409-425.

Kim, Illsoo. 1987. The Koreans: small business in an urban frontier. *New Immigrants in New York*. New York: Columbia University Press.

Light, Ivan. 1972. *Ethnic Enterprise in America*. Berkeley, CA: University of California Press.

Light, Ivan, and Edna Bonacich. 1991. *Immigrant Entrepreneurs: Koreans in Los Angeles, 1965-1982*. Berkeley, CA: University of California Press.

Light, Ivan, and Steven Gold. 2000. *Ethnic Economies*. Bingley, UK: Emerald Group.

Min, Pyong Gap. 1988a. *Minority Business Enterprise: Korean Small Business in Atlanta*. Staten Island, NY: The Center for Migration Studies.

Min, Pyong Gap. 1988b. Korean immigrant entrepreneurship: A multivariate analysis. *Journal of Urban Affairs* 10: 197-212.

Min, Pyong Gap. 1988c. *Ethnic Business and Economic Mobility: Korean Immigrants in Los Angeles*. Prepared for Presentation at the Annual Meeting of the American Sociological Association, Atlanta.

Min, Pyong Gap. 1990. Problems of Korean immigrant entrepreneurs. *The International Migration Review* 24(3): 436-455.

Morris, Stephen. 1956. Indians in East Africa: A study in a plural society. *The British Journal of Sociology* 7: 194-211.

Portes, Alejandro. 1998. Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology* 24(1): 1-24.

Putnam, Robert D. 1993. The prosperous community. *The American Prospect* 4(13): 35-42.

Sharma, Babita. 2019. *The Corner Shop: Shopkeepers, the Sharmas and the Making of Modern Britain.* Hodder & Stoughton.

Siu, Paul. 1952. The sojourner. *American Journal of Sociology* 58(1): 34-44.

Waldinger, Roger, Howard Aldrich, and Robin Ward. 1990. *Ethnic Entrepreneurs: Immigrant Business in Industrial Societies*. London: Sage Publications, Inc.

Winder, R. Bayly. 1962. The Lebanese in West Africa. *Comparative Studies in Society and History* 4: 296-333.

Wong, Eugene. 1985. Asian American middleman minority theory: The framework of an American myth. *Journal of Ethnic Studies* 13(1): 51-88.