Unpacking Sources of Comparative Advantage: A Quantitative Approach

(Supplementary Appendix: Not for Publication)

Davin Chor*
Singapore Management University

July 2010

Abstract

This Supplementary Appendix provides a detailed documentation for the dataset that was constructed for the analysis in this paper, expanding upon the description in the data appendix that accompanies the published version. It also contains several Appendix Tables of summary statistics. The second part of the Supplementary Appendix reports details related to the execution of the simulated methods of moments (SMM) estimation, including some sensitivity analysis.

*School of Economics, Singapore Management University, 90 Stamford Rd, Singapore 178903. Tel: +(65) 6828-0876. Fax: +(65) 6828-0833. E-mail: davinchor@smu.edu.sg

1 Data Appendix

Appendix Table 1 lists the 83 countries and 20 industries in the dataset. Appendix Tables 2 and 3 report summary statistics for the country and industry variables respectively.

A. BILATERAL VARIABLES

Trade volumes: From Feenstra et al. (2005), for the year 1990, in thousands of current US dollars. Converted from SITC Rev 2 into US 1987 SIC format using a concordance based on Feenstra, Romalis and Schott (2002), henceforth FRS. FRS record US export data at the highly disaggregate Harmonized System (HS) 10-digit level, where each HS-10 product is also assigned a 5-digit SITC Rev 2 and a 4-digit SIC-87 category. This is used to derive concordance weights to map SITC Rev 2 categories into SIC-87 format, following the procedure in Cuñat and Melitz (2009).

Two complications arise. First, classification for the SIC-87 categories is based on observed finished products, but the distinction between SIC industries is often defined according to the production process. To cite an example from FRS, SIC 2011 and SIC 2013 are both for processed meats, with the difference being that 2011 conducts its own slaughtering while 2013 uses purchased carcasses. When products are observed at the dock, it is not possible to distinguish between the two, and so trade flows for both are merged under SIC 2011, with 2013 omitted from the FRS dataset. Table 1.3 in FRS lists the affected industries, detailing which categories have been excluded and which codes the export value has been merged under. I break up the merged trade flows for the affected categories in proportion to the value of US total shipments in 1990 reported in the NBER-CES database (Bartelsman et al. 2000).[1] Then, the SITC codes associated with the included SIC industry are also assigned to the previously excluded SIC industries. A second complication relates to Feenstra et al.'s (2005) use of SITC codes with suffixes 'A' and 'X', for trade flows not observed at a more disaggregate level. I assign the trade in these 'A' and 'X' categories to the truncated (more aggregate) SITC code. In other words, I treat 111A and 111X as coming from the 3-digit SITC category 111, and then use FRS to construct weights to map SITC 111 into SIC categories.

Trade flows were summed up to the 2-digit SIC level, yielding 20 industry groups. A zero is entered for all exporter-importer-industry cells for which no trade was reported.

Distance: Physical distance is measured by the great circle formula distance between countries' capital cities, taken from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). A country's log distance to itself is set to zero, so that physical distance does not impose an iceberg cost for internal trade. The following binary variables are also from the CEPII: (i) "Common Language", equal to 1 if at least 9% of each country's population speaks a shared language; and (ii) "Colony", equal to 1 if one of the countries had ever colonized the other. The

[1] One exception: SIC 2092 is excluded from FRS, with the associated trade flows being merged under SIC 0912 and 0913, which are primary fishing industries. Since shipment data for the 09XX categories is not available in the NBER-CES database, I imputed all of 0912 and 0913 to 2092.

"Border" dummy (equal to 1 if the countries share a land border) is coded using the CIA World Factbook. The following two measures are based on Rose (2004), augmented by direct reference to the WTO website to cover all country pairs in my sample for the year 1990: (i) "RTA", equal to 1 if the countries are joint signatories in any of the regional trade agreements reported to the WTO; and (ii) "GATT", equal to 1 if both countries are GATT/WTO members. A value of 1 is assigned for all five dummies for a country's distance from itself.

B. INDUSTRY CHARACTERISTICS

Factor intensities: From the NBER-CES database. Variables are calculated for each 2-digit SIC-87 industry. Skill intensity is the log of the ratio of non-production workers to total employment. Physical intensity is the log of the ratio of real capital stock to total employment. Both ratios are averages over the period 1980-89.

The welfare counterfactuals require information on industry factor payment shares. These are obtained from the same NBER-CES database, using averages over 1980-89. The share of payments to skilled labor (s_h) and unskilled labor (s_l) are calculated by the ratios of non-production worker payroll and production worker payroll to total industry value-added respectively. The factor share of physical capital (s_k) is the ratio of residual payments (total value-added minus total payroll) to total value-added.

External capital dependence ($CAPDEP$): Constructed following the methodology in Rajan and Zingales (1998). Data from Compustat is used, which covers all publicly-traded firms in North America. A given firm's dependence on external capital is the fraction of total capital expenditures over the period 1980-89 not financed by internal cash flow. The median value across firms in each SIC-87 2-digit category is used as the industry measure of $CAPDEP$. (The measure in Rajan and Zingales (1998) is constructed for a different classification system, namely ISIC 3- and 4-digit industries.)

Input concentration (HI): Constructed following Levchenko (2007). Equal to the Herfindahl index of intermediate input use, based on the 1987 US Input-Output (IO) Use Table. The IO-87 6-digit level categories map cleanly into the SIC-87 4-digit categories based on the correspondence table provided by the Bureau of Economic Analysis (BEA).[2] In the few cases where an IO-87 category maps into more than one SIC category, I split the inputs in proportion to US domestic shipments in the SIC destination categories, using the total shipments reported in the NBER-CES database as weights. Input use is then aggregated to the SIC 2-digit level, from which the input Herfindahl is calculated.

[2]Available at: http://www.bea.gov/bea/pn/ndn0016.zip. All SIC 4-digit industries are associated with a unique IO-87 6-digit category, except for SIC 3999 which is matched with two IO-87 6-digit categories.

Input Relationship-Specificity (RS): From Nunn (2007). RS is the share (by value) of inputs that are not sold on an organized exchange; this corresponds to the measure z^{rs2} in Nunn (2007). Data on input use is from the 1987 US Input-Output Use Table. Rauch (1999) provides the classification of goods into: (i) those sold on an organized exchange; (ii) those reference-priced in commercial publications; and (iii) goods that fall in neither of the above categories. Moving from (i) to (iii), one has successively more differentiated and hence more relationship-specific inputs. Rauch provides two codings, one "conservative" and one "liberal"; I use the "liberal" classification. I map the IO-87 codes to SIC-87 4-digit categories with the procedure described for the HI variable. The measure is aggregated up to the 2-digit level by taking a weighted average, using the share of total input consumption of each 4-digit industry as weights. (The measure in Nunn (2007) is constructed for IO-87 industries instead.)

Job complexity ($COMPL$): Based on Costinot (2009). The 1985 and 1993 instalments of the US Panel Survey of Income Dynamics (PSID) contain a question asking respondents to gauge how many months it would take a typical new employee with the requisite education background to become "fully trained and qualified" in the respondents' job. Costinot (2009) calculates the average response for SIC-1972 3-digit industries, normalized to a maximum value of 1. I assign these values to the corresponding 4-digit sub-categories. For missing 4-digit level observations, I assign the median complexity level observed at successively higher levels of industry aggregation (first at the 3-digit level, and if that is still missing, at the 2-digit level, and then at the 1-digit level). These are then transformed from SIC-1972 to SIC-1987 categories using the weights in the correspondence table developed by Bartelsman, Becker and Gray.[3] The value of $COMPL$ for each SIC-1987 2-digit industry is then taken as the median over all its 4-digit sub-categories. There are two industry groups for which this imputation procedure may seem too liberal, namely SIC 21 and 29, for which direct information on complexity is not available in the PSID for any of the 3-digit sub-categories. The OLS results are similar if I omit these two industry groups.

Sales Volatility ($SVOL$): From Cuñat and Melitz (2009). Equal to the employment-weighted standard deviation of sales growth for firms in the 1980-2004 Compustat sample. Only firms with at least 5 years of data are used. Observations where the absolute sales growth rate exceeds 300% are omitted as outliers.

C. COUNTRY CHARACTERISTICS

Factor endowments: Physical capital per worker ($\log(K/L)_i$) and human capital per worker ($\log(H/L)_i$) are from Hall and Jones (1999), for the year 1988.

[3] Available at: http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/Concordances/FromusSIC/sic7287.txt

Financial development ($FINDEV$): From Beck et al.'s (2000) Financial Structure and Economic Development Database, March 14 2005 update. Equal to the amount of credit extended by banks and other non-bank financial intermediaries to the private sector divided by GDP, averaged over 1980-89.

Legal System ($LEGAL$): From Gwartney and Lawson (2004). Index measure of "Legal System and Property Rights" for 1985, rescaled between 0 and 1, which is a composite of five sub-indices on: judicial independence; impartiality of courts; protection of intellectual property; military interference in the rule of law and the political process; and integrity of the legal system. These sub-indices are drawn from the International Country Risk Guide (ICRG) and the Global Competitiveness Report (GCR), the former being a private institutional assessment, while the latter is an international survey of business executives. The OLS results are similar using another popular index for the "rule of law" from the World Bank Governance Indicators (Kaufmann et al. 2005), available from 1996 onwards, which likely reflects the high persistence in institutional conditions over time in most countries.

Employment Flexibility ($FLEX$): From the World Bank's *Doing Business* database. Index of "Rigidity of Employment", averaged over 2003-06, rescaled to be increasing in labor market flexibility and to lie between 0 and 1. Calculated as the average of three sub-indices on: the difficulty of hiring a new worker; restrictions on expanding or contracting the number of working hours; and the difficulty and expense of dismissing a redundant worker. The indices are coded based on the methodology in Botero et al. (2004).

GDP: Both GDP and GDP per capita are taken from the World Development Indicators (WDI), in current US dollars.

Population: From the WDI.

Appendix Table 1A
List of SIC-87 2-digit Industries (20)

SIC Major groups: (2-digit level)

20: Food and Kindred Products
21: Tobacco Products
22: Textile Mill Products
23: Apparel and other Finished Products made from Fabrics and similar materials
24: Lumber and Wood Products, except Furniture
25: Furniture and Fixtures
26: Paper and Allied Products
27: Printing, Publishing, and Allied Industries
28: Chemicals and Allied Products
29: Petroleum Refining and Related Industries
30: Rubber and Miscellaneous Plastics Products
31: Leather and Leather Products
32: Stone, Clay, Glass, and Concrete Products
33: Primary Metal Industries
34: Fabricated Metal Products, except Machinery and Transportation Equipment
35: Industrial and Commercial Machinery, and Computer Equipment
36: Electronic and other Electrical Equipment, except Computer Equipment
37: Transportation Equipment
38: Measuring, Analyzing, and Controlling Instruments
(Photographic, Medical and Optical Goods; Watches and Clocks)
39: Miscellaneous Manufacturing Industries

Appendix Table 1B
List of Countries in Sample (83)

Countries: (ISO codes in parentheses)

Argentina (ARG); Australia (AUS); Austria (AUT); Burundi (BDI); Belgium (BEL); Bolivia (BOL); Brazil (BRA); Central African Republic (CAF); Canada (CAN); Switzerland (CHE); Chile (CHL); China (CHN); Ivory Coast (CIV); Cameroon (CMR); Colombia (COL); Costa Rica (CRI); Germany (DEU); Denmark (DNK); Dominican Republic (DOM); Algeria (DZA); Ecuador (ECU); Egypt (EGY); Spain (ESP); Finland (FIN); France (FRA); United Kingdom (GBR); Ghana (GHA); Greece (GRC); Guatemala (GTM); Honduras (HND); Haiti (HTI); Hungary (HUN); Indonesia (IDN); India (IND); Ireland (IRL); Iran (IRN); Israel (ISR); Italy (ITA); Jamaica (JAM); Jordan (JOR); Japan (JPN); Kenya (KEN); South Korea (KOR); Sri Lanka (LKA); Morocco (MAR); Madagascar (MDG); Mexico (MEX); Mali (MLI); Malawi (MWI); Malaysia (MYS); Niger (NER); Nigeria (NGA); Nicaragua (NIC); Netherlands (NLD); Norway (NOR); New Zealand (NZL); Pakistan (PAK); Panama (PAN); Peru (PER); Philippines (PHL); Papua New Guinea (PNG); Poland (POL); Portugal (PRT); Paraguay (PRY); Senegal (SEN); Singapore (SGP); Sierra Leone (SLE); El Salvador (SLV); Sweden (SWE); Syria (SYR); Chad (TCD); Togo (TGO); Thailand (THA); Tunisia (TUN); Turkey (TUR); Uganda (UGA); Uruguay (URY); United States (USA); Venezuela (VEN); South Africa (ZAF); Zaire (ZAR); Zambia (ZMB); Zimbabwe (ZWE)

Appendix Table 2A
Summary of Country Characteristics

	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
$\log(H/L)_i$	0.072	0.257	0.392	0.592	0.807	1.039	1.215	0.290
$\log(K/L)_i$	5.763	7.050	8.332	9.723	10.828	11.318	11.589	1.584
Financial Devt. ($FINDEV$)	0.007	0.100	0.157	0.279	0.515	0.790	1.378	0.296
Legal Quality ($LEGAL$)	0.17	0.26	0.35	0.5	0.67	0.79	0.83	0.185
Labor Mkt. Flexibility ($FLEX$)	0.225	0.39	0.49	0.615	0.76	0.87	1	0.184

Appendix Table 2B
Pairwise Correlation of Country Characteristics

	$\log(H/L)_i$	$\log(K/L)_i$	$FINDEV$	$LEGAL$
$\log(K/L)_i$	0.81***			
$FINDEV$	0.58***	0.66***		
$LEGAL$	0.69***	0.63***	0.68***	
$FLEX$	0.34***	0.28**	0.21*	0.23**

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Appendix Table 3A
Summary of Manufacturing Industry Characteristics
(20 industries, SIC-87 2-digit level)

	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
Skill intensity $(\log(H/L)^k)$	−1.971	−1.906	−1.576	−1.395	−1.082	−0.831	−0.759	0.370
Capital intensity $(\log(K/L)^k)$	2.316	2.891	3.499	3.906	4.589	5.071	6.127	0.884
Ext. Capital Dep. $(CAPDEP)$	−1.206	−0.751	−0.148	−0.028	0.165	0.587	0.941	0.498
Input Concentration (HI)	0.724	0.783	0.794	0.834	0.908	0.932	0.943	0.064
Input Relationship-Spec. (RS)	0.594	0.673	0.818	0.946	0.969	0.988	0.991	0.125
Job Complexity $(COMPL)$	0.148	0.153	0.311	0.384	0.615	0.732	1	0.221
Sales Volatility $(SVOL)$	0.124	0.130	0.144	0.152	0.179	0.198	0.219	0.026

Appendix Table 3B
Pairwise Correlation of Manufacturing Industry Characteristics
(20 industries, SIC-87 2-digit level)

	$\log(H/L)^k$	$\log(K/L)^k$	$CAPDEP$	HI	RS	$COMPL$
$\log(K/L)^k$	0.39*					
$CAPDEP$	0.52**	0.10				
HI	0.46**	−0.34	0.63***			
RS	0.09	−0.54**	0.13	0.55**		
$COMPL$	0.82***	0.19	0.65***	0.54**	0.21	
$SVOL$	−0.08	0.06	0.38*	0.11	−0.20	0.07

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

2 Further Details on the SMM Estimation

Appendix Table 4 lists the country and industry groups (with their associated cutoffs) used to reduce the number of fixed effects that need to be estimated in the SMM procedure. Appendix Table 5 reports the values of the data moments and the simulated moments evaluated at the SMM point estimates, based on a θ value of 8.28. There is clearly a good fit of the simulated moments to the actual data moments.

While we focus on the SMM estimates based on the central θ value of 8.28, some brief words about sensitivity are useful given the range of values of θ reported in this literature. We report in Appendix Table 6 the results from SMM estimation using $\theta = 3.60$ and $\theta = 12.86$, the lowest and highest values of θ respectively reported by Eaton and Kortum (2002). (Eaton and Kortum obtained these alternative values when applying different estimation strategies for θ.)

It turns out that lowering θ leads to implied values for β_d, β_f and β_{lm} that are larger in magnitude. For example, when θ is set equal to the lower range value of 3.60 from EK, I obtain an SMM estimate of -0.757 for the physical distance coefficient $\theta\beta_{d1}$ (significant at the 1% level), implying a β_{d1} value of $-0.757/3.60 = -0.21$. This is larger in magnitude than the value of $-0.919/8.28 = -0.11$ obtained when $\theta = 8.28$ is used instead. The converse happens when we raise θ. Intuitively, a lower θ raises the spread of the stochastic productivity shocks. Without a corresponding rise in the β_d, β_f and β_{lm} parameters that govern the strength of the systematic determinants of trade flows (distance, factor endowments, and institutions), the simulated trade patterns would be dominated by these stochastic shocks and ultimately yield a poor fit to the actual data. Nevertheless, note that while all the implied β parameters in Column 4a have increased in magnitude compared to Column 3a, their relative magnitudes to each other within the same Column remain fairly stable. In other words, the relative importance of each trade determinant remains similar to that from the baseline estimation with $\theta = 8.28$.

For the same reason, a similar increase in the magnitudes of β_d, β_f and β_{lm} is observed when the interval $[\underline{x}, \overline{x}]$ is expanded to cover more of the mass of the untruncated Gumbel distribution (results available on request). Once again, as the spread of the stochastic shocks increases, β_d, β_f and β_{lm} need to rise in magnitude to ensure that the simulated trade patterns are not driven entirely by the random shocks.

Appendix Table 4
List of Country and Industry Groups for SMM Estimation

Exporter Groups: (Grouped by total export volumes)

Group 1: TCD; MLI; TGO; BDI; MWI; CAF; NIC; UGA; SLE; NER; MDG; BOL; SYR; SLV; HTI; PRY; JOR; PNG; SEN; NGA; HND; GHA; ZWE; CMR; KEN; GTM ($< \text{US\$}1,000,000$)

Group 2: JAM; ZMB; IRN; CRI; ECU; CIV; LKA; EGY; URY; ZAR; PAN; DOM; PER; TUN; COL; PAK; MAR ($> \text{US\$}1,000,000$ and $< \text{US\$}5,000,000$)

Group 3: DZA; HUN; CHL; GRC; VEN; NZL; POL; PHL; ZAF; ARG; TUR; ISR ($> \text{US\$}5,000,000$ and $< \text{US\$}10,000,000$)

Group 4: IND; IDN; PRT; NOR; THA; AUS; IRL; FIN; MYS; BRA; MEX; DNK; SGP; AUT; CHN; ESP; KOR; SWE; CHE ($> \text{US\$}10,000,000$ and $< \text{US\$}90,000,000$)

Group 5: CAN; BEL; NLD; GBR; ITA; FRA; JPN; USA; DEU ($> \text{US\$}90,000,000$)

SIC Industry Groups: (Grouped by total trade volumes)

Group 1: 21; 27; 25; 31; 32; 24 ($< \text{US\$}50,000,000$)

Group 2: 30; 29; 22; 39; 26; 23; 34; 38; 33; 20 ($> \text{US\$}50,000,000$ and $< \text{US\$}200,000,000$)

Group 3: 36; 28; 37; 35 ($> \text{US\$}200,000,000$)

Appendix Table 5
Comparison of Data Moments and Matched Simulated Moments ($\theta = 8.28$)

	Data moment	Matched moment (based on $\hat{\Theta}^{SMM}$)
Regression coefficients:		
β_{d1}: Log (Distance)	−1.16085	−1.16066
β_{d2}: Common Language	0.50195	0.50212
β_{d3}: Colony	0.76554	0.76529
β_{d4}: Border	0.18948	0.18979
β_{d5}: RTA	0.29025	0.29092
β_{d6}: GATT	0.21723	0.21572
β_{f1}: $\log(H/L)^k \times \log(H/L)_i$	1.24570	1.24543
β_{f2}: $\log(K/L)^k \times \log(K/L)_i$	0.16413	0.16408
β_{lm1}: $CAPDEP \times FINDEV$	1.27883	1.27978
β_{lm2}: $HI \times LEGAL$	14.30727	14.30744
β_{lm3}: $RS \times LEGAL$	9.63769	9.63810
β_{lm4}: $COMPL \times LEGAL$	2.91853	2.91879
β_{lm5}: $COMPL \times \log(H/L)$	1.46171	1.46231
β_{lm6}: $SVOL \times FLEX$	9.04316	9.04320
Trade shares:		
Exporter Group 2:	0.01398	0.01400
Exporter Group 3	0.03829	0.03833
Exporter Group 4	0.23768	0.23812
Exporter Group 5	0.70615	0.70660
SIC Group 2	0.41362	0.41393
SIC Group 3	0.51905	0.51933

Appendix Table 6
SMM Estimates: Sensitivity to θ

	(3) SMM $\theta = 8.28$	(3a) Implied β's $\theta = 8.28$	(4) SMM $\theta = 3.60$	(4a) Implied β's $\theta = 3.60$	(5) SMM $\theta = 12.86$	(5a) Implied β's $\theta = 12.86$
Distance and Geography:						
$\theta\beta_{d1}$: Log (Distance)	−0.919*** (0.002)	−0.111	−0.757*** (0.001)	−0.210	−0.965*** (0.003)	−0.075
$\theta\beta_{d2}$: Common Language	0.400*** (0.002)	0.048	0.332*** (0.001)	0.092	0.406*** (0.005)	0.032
$\theta\beta_{d3}$: Colony	0.603*** (0.003)	0.073	0.501*** (0.001)	0.139	0.618*** (0.006)	0.048
$\theta\beta_{d4}$: Border	0.130*** (0.003)	0.016	0.106*** (0.001)	0.029	0.140*** (0.006)	0.011
$\theta\beta_{d5}$: RTA	0.192*** (0.003)	0.023	0.196*** (0.001)	0.054	0.202*** (0.005)	0.016
$\theta\beta_{d6}$: GATT	0.168*** (0.008)	0.020	0.130*** (0.003)	0.036	0.144*** (0.015)	0.011
Heckscher-Ohlin: (industry char. × country char.)						
$\theta\beta_{f1}$: $\log(H/L)^k \times \log(H/L)_i$	1.245*** (0.037)	0.150	0.920*** (0.015)	0.256	1.344*** (0.053)	0.104
$\theta\beta_{f2}$: $\log(K/L)^k \times \log(K/L)_i$	0.093*** (0.002)	0.011	0.084*** (0.001)	0.023	0.098*** (0.003)	0.008
Institutional: (industry char. × country char.)						
$\theta\beta_{lm1}$: $CAPDEP \times FINDEV$	0.883*** (0.011)	0.107	0.738*** (0.006)	0.205	0.920*** (0.020)	0.072
$\theta\beta_{lm2}$: $HI \times LEGAL$	8.867*** (0.341)	1.071	8.359*** (0.161)	2.322	8.842*** (0.630)	0.688
$\theta\beta_{lm3}$: $RS \times LEGAL$	7.032*** (0.143)	0.849	5.902*** (0.078)	1.639	7.258*** (0.222)	0.564
$\theta\beta_{lm4}$: $COMPL \times LEGAL$	3.426*** (0.154)	0.414	3.456*** (0.063)	0.960	3.322*** (0.231)	0.258
$\theta\beta_{lm5}$: $COMPL \times \log(H/L)_i$	0.611*** (0.078)	0.074	0.391*** (0.033)	0.109	0.699*** (0.112)	0.054
$\theta\beta_{lm6}$: $SVOL \times FLEX$	8.831*** (0.381)	1.067	8.190*** (0.167)	2.275	10.070*** (0.582)	0.783

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively. Columns 3, 4 and 5 present the SMM coefficients obtained when different θ values are used. In particular, Column 3 reproduces the results from Table 2 for the central case of $\theta = 8.28$. The exporter and industry fixed effects have been grouped (as discussed in the text); these group fixed effects are not reported. Columns 3a, 4a and 5a present the respective implied values for the β_d, β_f and β_{lm} parameters, after dividing the SMM coefficients by θ.