Journal of International Economics 73 (2007) 278–308
www.elsevier.com/locate/econbase

A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises

Jeffrey H. Bergstrand [a,b,*], Peter Egger [c,d,e]

[a] *Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, Notre Dame, IN 46556, USA*
[b] *CESifo, Munich, Germany*
[c] *Department of Economics, University of Munich, Munich, Germany*
[d] *IFO Institute, Munich, Germany*
[e] *CESifo, Munich, Germany*

Received 8 March 2006; received in revised form 25 October 2006; accepted 22 March 2007

Abstract

This paper addresses two important issues at the nexus of the literatures on international trade, foreign direct investment (FDI), foreign affiliate sales (FAS), and multinational enterprises (MNEs). First, the introduction of a third internationally-mobile factor (physical capital) to the standard 2×2×2 "knowledge-capital" model of MNEs with skilled and unskilled labor allows us to resolve fairly readily the puzzle in the modern MNE literature that foreign affiliate sales among two identical economies completely displace their international trade. Intra-industry trade *and* intra-industry FDI (and FAS) can coexist for national and multinational firms (with identical productivities) in identical countries. Second, the introduction also of a third country to the model suggests a formal *N*-country theoretical rationale for estimating gravity equations of bilateral FDI flows and FAS, in a manner consistent with estimating gravity equations for bilateral trade flows.

Keywords: Foreign direct investment; International trade; Multinational firms; Gravity equation

JEL classification: F11; F12; F21; F23

* Corresponding author. Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, IN 46556, USA. Tel.: +1 574 631 6761; fax: +1 574 631 5255.
E-mail address: Bergstrand.1@nd.edu (J.H. Bergstrand).

0022-1996/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jinteco.2007.03.004

1. Introduction

> "Multinationals displace trade…." (Markusen, 1995, p. 180).

> "The cross-country pattern of FDI is quite well approximated by the 'gravity' relationship." (Barba Navaretti and Venables, 2004, p. 32).

This paper addresses two important issues at the nexus of the literatures on international trade, foreign direct investment (FDI), foreign affiliate sales (FAS), and multinational enterprises (MNEs). First, the modern $2\times2\times2$ general equilibrium theory of MNEs synthesized in Markusen (2002) implies that – in two countries with identical absolute and relative factor-endowments (other things equal) – horizontal MNEs' foreign affiliate sales displace *completely* national firms (with identical productivities) and trade between the two countries. However, the European Union and the United States, for instance, have both the largest intra-industry bilateral foreign direct investment flows (and FAS) *and* intra-industry trade flows. This is a puzzle. Second, while multi-country theoretical foundations for the trade gravity equation are now well established (cf., Anderson and van Wincoop, 2004, and Feenstra, 2004, ch. 5, for overviews), there have been virtually no formal N-country ($N>2$) theoretical frameworks provided in international economics for estimating gravity equations of aggregate bilateral FDI, despite numerous empirical studies over the past decade using the gravity equation to explain such flows.[1] Blonigen et al. (in press) note that the "gravity model is arguably the most widely used empirical specification for FDI" (p. 8). Yet the typical rationale for applying the gravity equation to bilateral FDI is by analogy to the trade gravity equation, cf., Mutti and Grubert (2004, p. 339) and Blonigen (2005, p. 21). This suggests a puzzle similar to one posed 30 years ago for trade: The gravity equation explains bilateral FDI empirically quite well… *but why*?

We suggest a simple, integrated solution to both puzzles. First, the introduction of a third factor – physical capital – to Markusen's two-factor "knowledge-capital" model with only skilled and unskilled labor, combined with the assumption that headquarters (plant) setups require human (physical) capital, implies that national exporting enterprises (NEs) can coexist with horizontal MNEs (HMNEs) in pairs of countries with identical relative *and* absolute factor-endowments (and all firms sharing identical technologies). With skilled labor not being the only factor used to setup both plants and firms, skilled labor is not completely displaced from plant setups to firm setups as two countries' GDPs converge in size. In the $2\times2\times2$ model of Markusen and Venables (2000), there is only a (highly unlikely) unique combination of trade costs, investment costs and ratio of

[1] Martin and Rey (2004) have advanced a theory of the gravity equation for bilateral *portfolio* investment flows. However, no studies have developed a theory for the gravity equation for bilateral FDI stocks/flows. Recently, a few studies have offered theoretical rationales for estimating bilateral FAS gravity equations, cf., Grazalian and Furtan (2005), Kleinert and Toubal (2005), and Lai and Zhu (2006); however, these studies assume exogenously heterogeneous productivities to generate coexistence of multinational and national firms, in the spirit of the model in Helpman et al. (2004), and only explain FAS. By contrast, we motivate the endogenous coexistence of country pairs' horizontal MNEs and national firms sharing identical productivities, even for identical economies. Moreover, by incorporating capital flows, we provide simultaneously theoretical rationales for trade, FAS, *and* FDI gravity equations, distinguishing explicitly between FDI (a measure of MNE capital flow) and FAS (a measure of MNE production). Consequently, our analysis is in the spirit of the "new trade theory" where, as noted in Helpman (2006, p. 592), within-industry heterogeneity results from product differentiation and monopolistic competition, and heterogeneous productivities among firms are unnecessary to explain large volumes of intra-industry trade. Since our theoretical model will be static, conventional to the trade and knowledge–capital MNE literatures (cf., Markusen, 2002), bilateral "flows" and "stocks" are conceptually identical. In empirical work, however, flows and stocks will be distinguished. Traditionally, gravity equations have used bilateral FDI stock data, cf., Blonigen et al. (in press).

plant-to-firm setup costs where NEs and HMNEs can coexist. By contrast, with three factors, NEs and HMNEs with identical productivities can coexist over a wide range of combinations of trade costs, investment costs, and plant-to-firm setup costs due to the endogenous adjustment of the relative price of human-to-physical capital. Intra-industry trade *and* FDI can coexist for two *identical* countries.[2] Moreover, the existence of human and physical capital allows distinguishing MNEs' foreign affiliate production from MNEs' capital flows, two different "concepts of FDI."[3]

Second, while introducing a third factor implies coexistence of NEs and HMNEs for identically-sized economies, this extension cannot explain empirically the "complementarity" of bilateral trade and FDI flows to GDP similarity. Typical empirical gravity equations of international trade and FDI tend to suggest that both trade and FDI from country *i* to country *j* should be positively related to the size *and* similarity of their GDPs. However, the introduction of a third country – rest-of-world (ROW) – to our "knowledge-and-physical-capital" model can explain readily the complementarity of bilateral trade, FDI, and FAS to changes in a pair of countries' economic size and similarity as typical to gravity equations. In a two-country world, gross multilateral and bilateral trades (or FAS) are identical; NEs and HMNEs must substitute for one another when the two countries are identically sized in the face of trade and investment costs. However, introducing a third country along with capital mobility allows two countries' trade, FDI, and FAS to covary positively with increases in these two countries' GDP similarity because the "substitution effect" associated with exogenous trade-to-investment costs is potentially offset by a "complementarity effect" generated by endogenous relative prices of physical-to-human capital interacting with the three countries' economic sizes. With three countries, both bilateral trade *and* FAS are maximized when a pair of countries' GDPs is identical, unlike a two-country world. Moreover, the presence of the third country can explain why FDI from one country to another is *not* maximized when GDPs are identical — which is observed empirically!

The complexity of the nonlinear relationships between trade, foreign affiliate sales, FDI, GDP sizes and similarities, trade costs, and investment costs precludes closed-form solutions, such as the trade gravity equations derived in Helpman and Krugman (1985) or Anderson and van Wincoop (2003). As in Markusen (2002), we must employ a numerical version of the general equilibrium (GE) model to motivate the gravity-like relationships among these variables. However, the theoretical relationships between the bilateral flows – trade, foreign affiliate sales, and FDI – and GDP sizes and similarities demonstrated graphically are similar, but not identical. Consequently, to avoid "ocular-metrics" and in the presence of approximation errors associated with the nonlinear relationships, we employ regressions of "theoretical" flows on pairs of countries' GDPs, trade and investment costs, and ROW GDP to motivate theoretically *and*

[2] We are not the first to try to address this puzzle. Blonigen (2001) and Head and Ries (2001a,b) introduce intermediate goods to address the issue. However, the nature of trade in these models is "vertical," nor horizontal; Baldwin and Ottaviano (2001) argue that the "intermediates augmented proximity-versus-scale model still has trade and FDI as substitutes" (p. 432). As an alternative, Baldwin and Ottaviano argue that obstacles to trade generate an incentive for multiproduct firms to simultaneously engage in FDI and trade, in the spirit of the Brander–Krugman "reciprocal-dumping" model for goods trade. Our focus on a third factor and third country is an alternative to this.

[3] See Lipsey (2002) for more on the distinction between FAS and FDI. The modern $2\times2\times2$ MNE model explains trade and FAS simultaneously, but not FDI. As Markusen (2002) notes, "the models in this book are addressed more closely to affiliate output and sales than to investment stocks" (p. 8). Similarly, the heterogeneous productivities models cited in Footnote 1 cannot explain FDI as these models have only one factor, immobile labor.

quantitatively that bilateral trade, FAS, and FDI flows' economic determinants should be "well-approximated" by gravity equations — yet not precisely the *same* gravity relationships.[4]

The remainder of the paper is as follows. Section 2 presents some typical empirical regression results for trade and FDI gravity equations to motivate our analysis. Section 3 presents our $3 \times 3 \times 2$ GE model of FDI, MNEs and trade. Section 4 discusses the calibration of the numerical version of the GE model. Section 5 shows how introduction of the third factor generates "coexistence" of NEs (bilateral intra-industry trade) and HMNEs (bilateral intra-industry investment) for two identical countries. Section 6 demonstrates how the further introduction of a third country generates "complementarity" of bilateral FAS, FDI, and trade flows with respect to the country pair's economic size and similarity. Section 7 addresses the regressions of "theoretical flows" on GDP sizes and similarities, trade and investment costs, and ROW GDP to show that bilateral trade, FAS, and FDI flows' determinants should be "well-approximated" by gravity equations — though not quantitatively identical gravity relationships. Section 8 concludes.

2. Typical empirical gravity equations for trade and FDI flows

In a simple theoretical world of N (>2) countries, one final differentiated good, no trade costs, but internationally-immobile factors (e.g., labor and/or capital), we know from the trade gravity–equation literature that the trade flow from country i to country j in year t (Flow_{ijt}) will be determined by:

$$\text{Flow}_{ijt} = \text{GDP}_{it}\text{GDP}_{jt}/\text{GDP}_t^W \tag{1}$$

where GDP^W is world GDP, or in log-linear form:

$$\ln\,\text{Flow}_{ijt} = -\ln(\text{GDP}_t^W) + \ln(\text{GDP}_{it}) + \ln(\text{GDP}_{jt}). \tag{2}$$

Of course, the real world is not so generous as to allow international trade to be "frictionless." Hence, specification (2) is augmented typically to include various empirical proxies for bilateral trade costs, such as time-invariant bilateral distance (DIST_{ij}), a dummy variable for common language (LANG_{ij}), a (potentially time-varying) dummy variable for common membership in a regional trade agreement (RTA_{ijt}), and a time trend. Table 1 provides representative gravity equations using pooled cross-section time-series empirical data for bilateral trade and FDI flows among the 17 most developed OECD countries for 11 years (1990–2000) in columns (1)–(4).[5]

[4] We will not provide a theoretical framework that is the perfect complement to the Anderson and van Wincoop (2003) theoretical foundation using a "conditional" general equilibrium (GE) model, which enabled them to derive a closed-form, analytic solution for the trade gravity equation with unit GDP elasticities. Instead, we use an "unconditional" GE approach, where the allocation of bilateral flows across countries is non-separable from production and consumption allocations within countries, whereas a "conditional" GE approach assumes separability of bilateral flows from production and consumption allocations, cf., Anderson and van Wincoop (2004). Our model is in the spirit of Markusen (2002). The benefit of our approach is illuminating the endogenous determination of representative firms' outputs, countries' numbers of NEs vs. MNEs, allocation of MNEs' capital between countries, etc. The cost is we forego a closed-form solution for the gravity equation, relying instead on regressions of "theoretical" flows (from a numerical GE model) on GDPs, etc. to motivate theoretically estimation of log-linear gravity equations for trade flows, FAS, and FDI flows. Shim (2005) extends the Anderson and van Wincoop conditional GE model to include FAS to further explain the "border puzzle"; however, Shim assumes exogenously the (unlikely) unique equilibrium in Markusen and Venables (2000) that generates coexistence of NEs and MNEs.

[5] For our sample, RTA has cross-sectional but no time variation, as the Canadian–U.S. FTA and the EC, EFTA, and EC-EFTA FTAs were in place over the entire period. The limitation on number of countries and years in the sample was dictated by measures of bilateral trade and investment costs to be addressed later.

While numerous studies separately have estimated gravity equations for trade flows and for FDI flows, few studies have estimated such equations for both flows using the same specification and years.[6] Specification (1) provides empirical results for trade using the specification analogous to Eq. (2) including typical RHS covariates. Coefficient estimates have plausible values, consistent with the literature.

Of course, the standard frictionless gravity equation can be altered algebraically to separate economic size ($\mathrm{GDP}_i+\mathrm{GDP}_j$) and similarity ($s_i s_j$), where $s_i=\mathrm{GDP}_i/(\mathrm{GDP}_i+\mathrm{GDP}_j)$ and analogously for j:

$$\mathrm{Flow}_{ijt}=\mathrm{GDP}_{it}\mathrm{GDP}_{jt}/\mathrm{GDP}_t^W=(\mathrm{GDP}_{it}+\mathrm{GDP}_{jt})^2(s_{it}s_{jt})/\mathrm{GDP}_t^W. \tag{3}$$

When countries i and j are identical in size ($s_i=s_j=1/2$), $s_i s_j$ is at a maximum. In log-linear form, (3) is:

$$\ln\mathrm{Flow}_{ij}=-\ln(\mathrm{GDP}^W)+2\ln(\mathrm{GDP}_i+\mathrm{GDP}_j)+\ln(s_i s_j). \tag{4}$$

Specification (2) in Table 1 for trade flows provides results using the formulation in Eq. (4). Specifications (3) and (4) in Table 1 provide coefficient estimates for representative FDI gravity equations for the same OECD countries and years. We note three interesting results.

First, the gravity equation works almost as well for bilateral FDI as for bilateral trade (using R^2 values), as typically found. Bilateral trade and FDI both increase in economic size *and* similarity. These results suggest that NEs and HMNEs coexist when pairs of (similar per capita income) countries have identical GDPs, and that trade and FDI are complements with respect to GDP similarity. This conflicts with the modern $2\times2\times2$ theory of MNEs.

Second, we observe a notable asymmetry from comparing specification (1) with (3). For trade, the exporter's GDP elasticity is similar to the importer's. However, for FDI the exporter's GDP elasticity is significantly greater than the importer's. Alternatively, a comparison of specification (2) with (4) reveals notably different elasticities of trade and FDI with respect to GDP similarity.[7] These results suggest that bilateral trade and FDI are "well-approximated by the gravity relationship" — but not precisely the *same* gravity equation. This is a puzzle.

Third, following Egger (2000) and others, in the presence of panel data, it has become common to include bilateral-pair fixed effects to eliminate any omitted variables bias associated with unobserved time-invariant pair-specific heterogeneity (not captured by bilateral distance and the language and RTA dummies). Columns (5)–(8) report the results for estimating the trade and FDI gravity equations using pair-specific fixed effects. Since only GDPs have time variation, their coefficient estimates are presented using within variation (Within-R^2 values reported). Specifications (5)–(8) provide even *more striking* asymmetries in GDP coefficient estimates. The notable difference between specifications (5) and (7) is that the exporting (home FDI) country's GDP elasticity is much smaller (larger) than the importing (host FDI) country's GDP elasticity. For specifications (6) and (8), as before for specifications (2) and (4), the GDP

[6] Bilateral FAS data is currently available only for a few developed countries and so omitted. By contrast, trade and FDI data is available for many countries. A time trend is included. Data are described in a Data Appendix.

[7] Bilateral distance and language-dummy coefficient estimates are also quantitatively different; however, this issue is outside our scope. The different coefficient estimates for the RTA dummy variable are addressed later.

Table 1
Coefficient estimates from gravity equations*

RHS variables	Empirical (LHS) flows, pooled cross-section, time-series data				Empirical (LHS) flows, bilateral fixed effects				"Theoretical" flows for LHS variables				Empirical (LHS) flows, bilateral fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	ln Trade_{ijt}	ln Trade_{ijt}	ln FDI_{ijt}	ln FDI_{ijt}	ln Trade_{ijt}	ln Trade_{ijt}	ln FDI_{ijt}	ln FDI_{ijt}	ln Trade_{ijt}	ln Trade_{ijt}	ln FDI_{ijt}	ln FDI_{ijt}	ln Trade_{ijt}	ln Trade_{ijt}	ln FDI_{ijt}	ln FDI_{ijt}
ln GDP_{it}	0.82		0.93		0.52		2.01		0.35		0.59		0.29		1.47	
	(86.77)		(39.62)		(5.44)		(5.74)		(89.13)		(36.00)		(2.19)		(12.85)	
ln GDP_{jt}	0.80		0.78		1.51		1.39		0.58		0.21		1.04		0.73	
	(85.62)		(33.83)		(15.71)		(4.02)		(146.80)		(146.80)		(7.80)		(1.38)	
ln GDPT_{ijt}		1.62		1.78		2.07		3.45		0.99		0.95		1.36		2.11
		(111.93)		(50.45)		(14.58)		(6.75)		(138.25)		(61.99)		(5.85)		(2.17)
ln GDPSim_{ijt}		0.81		1.28		0.64		1.31		0.74		1.53		0.61		1.18
		(38.02)		(25.86)		(5.07)		(2.95)		(83.50)		(75.13)		(4.13)		(2.09)
ln DIST_{ij}	−0.62	−0.62	−0.84	−0.75												
	(−42.43)	(−41.09)	(−24.24)	(−21.11)												
ln LANG_{ij}	0.67	0.67	1.32	1.36												
	(19.69)	(19.65)	(16.70)	(17.42)												
RTA_{ijt}	0.57	0.57	−0.53	−0.47												
	(14.52)	(14.48)	(−5.79)	(−5.17)												
ln ROWGDP_{ijt}													−4.10	−3.80	−9.18	−9.75
													(−4.86)	(−3.42)	(−2.90)	(2.20)
ln transport costs_{ijt}													−0.02	−0.02	−0.003	−0.003
													(−5.78)	(−5.73)	(−0.21)	(−0.21)
ln investment costs_{ijt}													−0.004	−0.005	−0.01	−0.01
													(−2.73)	(−3.66)	(−1.94)	(−1.72)
ln $[(S/U)_{it}/(S/U)_{jt}]$													1.81	1.49	5.48	5.73
													(5.35)	(4.47)	(4.18)	(4.43)
ln $[(K/U)_{it}/(K/U)_{jt}]$													−0.07	−0.20	0.56	0.67
													(−0.09)	(−2.65)	(1.73)	(2.21)
R^2	0.87	0.87	0.57	0.58	0.49	0.49	0.33	0.33	0.89	0.86	0.37	0.71	0.51	0.49	0.28	0.27
No. of pairs	272	272	266	266	272	272	266	266	272	272	266	266	263	263	254	254
No. of observations	2983	2983	2683	2683	2983	2983	2683	2683	2983	2983	2683	2683	2069	2069	1828	1828

*T−statistics are in parentheses. In empirical regressions, time dummies are included, but coefficient estimates omitted. Bilateral fixed effects' coefficient estimates are also not reported for brevity. In bilateral fixed effects (empirical) regressions, R^2 denotes "Within-R^2." All RHS variables use empirical data.

similarity elasticity is notably smaller for trade than for FDI flows. We address this puzzle as well.[8]

3. Theoretical framework

The model we use is a straightforward extension of the $2\times2\times2$ knowledge-capital model in Markusen (2002) with NEs, HMNEs, and vertical MNEs (VMNEs). The demand side is modeled analogously to that model. However, we extend that model in two ways. The first distinction is to use *three* primary factors of production: unskilled labor, skilled labor (or human/knowledge capital), and physical capital. We assume unskilled and skilled labor are immobile internationally, but physical capital is mobile in the sense that MNEs will endogenously choose the optimal allocation of domestic physical capital between home and foreign locations to maximize profits, consistent with the BEA definition of foreign "direct investment positions" using domestic and foreign-affiliate shares of real fixed investment.[9] The introduction of a third factor – combined with an assumption that the setup of a headquarters requires home skilled labor (to represent, say, research and development) while the setup of a plant in any country requires the home country's physical capital (to represent, say, equipment) – helps explain "coexistence" of HMNEs and NEs for two identically-sized developed-countries for a wide range of parameter values. Differentiated final goods are produced from physical capital, skilled labor, and unskilled labor.[10]

The second distinction of our model is to introduce a third country. The presence of the third country helps to explain the "complementarity" of bilateral FAS and trade with respect to a country pair's economic size and similarity and that bilateral FDI empirically tends to be maximized when the home country's GDP is larger than the host country's. One implication of a third country is that (in equilibrium) both two-country HMNEs and three-country HMNEs may surface. As a first venture into the 3-factor, 3-country setting, we limit our scope in this paper to the 3-country HMNE equilibrium and to studying inter-developed-country bilateral interactions, but where ROW can be either a developed or developing economy.[11] Due to space constraints, we leave study of other interactions for future work.

[8] See Egger (2000) for more on the appropriateness of bilateral-pair-specific fixed effects. As noted in Anderson and van Wincoop (2003), an important potential source of omitted variables bias in gravity equations is the role of "multilateral resistance" terms. In the context of a panel, this suggests including also country-specific dummies for each time period; these "country-and-time" effects are addressed in Baier and Bergstrand (2007). For our sample of 17 OECD countries for 1990–2000, we assume here that the MR terms are slow moving, so bilateral pair "fixed" effects should capture the (most important) cross-sectional influence of these terms and time-invariant pair-specific unobserved heterogeneity. This allows estimating coefficients of the GDP terms using time variation.

[9] In the typical $2\times2\times2$ model, headquarters use home skilled labor exclusively for setups; home (foreign) plants use home (foreign) skilled labor for setups (cf., Markusen, 2002, p. 80). With only immobile skilled and unskilled labor, the 2-factor models preclude home physical capital being utilized to setup foreign plants. We often refer to the transfer of physical capital by MNEs as capital "mobility," in the tradition of the pre-1960 literature, cf., Mundell (1957, pp. 321–323) and Lipsey (2002); practically speaking, it is easiest to think of FDI as "greenfield" investment. Consistent with Markusen (2002) and the modern MNE literature, the model is real; there are no paper assets. Moreover, while physical capital can be utilized in different countries, ownership of any country's endowment of such capital is immobile, again following Mundell (1957). In reality, the presence of (paper) claims to physical capital allows much easier transfer of resources and is *one* way of measuring FDI. However, the "current-cost" method of measuring FDI is related to the shares of an MNE's *real* fixed investment in plant and equipment that is allocated to the home country relative to foreign affiliate(s); this effectively measures physical capital mobility, cf., Borga and Yorgason (2002, p. 27).

[10] As in Markusen (2002), internationally-immobile skilled labor still creates firm-specific intangible assets that are costlessly shared internationally by MNEs with their plants. This aspect is maintained.

[11] In earlier versions (cf., Bergstrand and Egger, 2004, 2006), we included notation explicitly for possible 2-country HMNEs. However, space constraints limit discussion here to only the 3-country HMNEs, noting that all the results generalize to allowing 2-country HMNEs.

3.1. Consumers

Consumers are assumed to have a Cobb–Douglas utility function between final differentiated goods (X) and homogeneous goods (Y). Consumers' tastes for differentiated products (e.g., manufactures) are assumed to be of the Dixit–Stiglitz constant elasticity of substitution (CES) type, as typical in trade. We let V_i denote the utility of the representative consumer in country i (i=1, 2, 3). Let η be the Cobb–Douglas parameter reflecting the relative importance of manufactures in utility and ε be the parameter determining the constant elasticity of substitution, σ, among these manufactured products ($\sigma \equiv 1-\varepsilon$, $\varepsilon<0$). Manufactures can be produced by three different firm types: national firms (n), horizontal multinational firms (h), and vertical multinational firms (v). In equilibrium, some of these firms may not exist (depending upon absolute and relative factor-endowments and parameter values). These will be reflected in three sets of components in the first of two RHS bracketed terms in Eq. (5) below:

$$V_i = \left[\sum_{j=1}^{3} n_j \left(\frac{x_{ji}^n}{t_{Xji}}\right)^{\frac{\varepsilon}{\varepsilon-1}} + \sum_{j=1}^{3} h_j (x_{ii}^h)^{\frac{\varepsilon}{\varepsilon-1}} + \sum_{k\neq j}\sum_{j=1}^{3} v_{kj}\left(\frac{x_{ji}^v}{t_{Xji}}\right)^{\frac{\varepsilon}{\varepsilon-1}}\right]^{\frac{\varepsilon-1}{\varepsilon}\eta} \left[\sum_{j=1}^{3} Y_{ji}\right]^{1-\eta}. \tag{5}$$

The first component reflects *national* (non-MNE) firms that can produce differentiated goods for the home market or export to foreign markets from a single plant in the country with its headquarters, where x_{ji}^n denotes the (endogenous) output of country j's national firm in industry X sold to country i, n_j is the (endogenous) number of these national firms in j, and t_{Xji} is the gross trade cost of exporting X from j to i.

The second component reflects *horizontal* MNEs that have plants in foreign countries to be "proximate" to markets to avoid trade costs; horizontal MNEs cannot export goods. In a three-country HMNE equilibrium, every HMNE has a plant in its headquarters country and two foreign plants. Let x_{ii}^h denote the output of a representative (three-country) HMNE producing in i and selling in i and h_j denote the number of multinationals that produce in all (three) countries and are headquartered in j.

The third component reflects *vertical* MNEs. VMNEs have a headquarters in one country and a plant in one of the other countries, just not in the headquarters country. The primary motivation for a VMNE is "cost differences"; different relative factor intensities and relative factor abundances motivate separating headquarters and production into different countries. Let v_{kj} denote the number of VMNEs with headquarters in k and a plant in j ($j \neq k$) with the plant's output potentially sold to any country (including k); such VMNEs include global "export-platform" MNEs, such as discussed in Ekholm, Forslid, and Markusen (in press) and Blonigen et al. (in press). Let x_{ji}^v denote the output of the representative VMNE with production in j and consumption in i.

In the second bracketed RHS term, let Y_{ji} denote the output of the homogenous (e.g., agriculture) good produced in country j under constant returns to scale using unskilled labor and consumed in i.

We let t_{Xji} (t_{Yji}) denote the gross trade cost for shipping differentiated (homogeneous) good X (Y) from j to i.[12] Let t_{Xji}=1 for i=j, and analogously for t_{Yji}. It will be useful to define:

$$\begin{aligned} t_{Xji} &= (1+b_{Xji})(1+\tau_{Xji}) \\ t_{Yji} &= (1+b_{Yji})(1+\tau_{Yji}) \end{aligned}$$

[12] For modeling convenience, we define Y_{ji} net of trade costs; t_{Yji} surface explicitly in the factor-endowment constraints.

where τ denotes a "natural" trade cost of physical shipment (cif/fob−1) of the "iceberg" type, while b represents a "policy" trade cost (i.e., tariff rate) which generates potential revenue. For instance, b_{Xji} denotes the tariff rate (e.g., 0.05=5%) on imports from j to i in differentiated final good X.

The budget constraint of the representative consumer in country i is assumed to be:

$$\sum_{j=1}^{3} n_j p_{Xj}^{n} x_{ji}^{n} + \sum_{j=1}^{3} h_j p_{Xi}^{h} x_{ii}^{h} + \sum_{k\neq j}\sum_{j\neq i} v_{kj} p_{Xj}^{v} x_{ji}^{v} + \sum_{j=1}^{3} p_{Yj} Y_{ji}$$
$$= r_i K_i + w_{Si} S_i + w_{Ui} Ui + \sum_{j\neq i} n_j b_{Xji} p_{Xj}^{n} x_{ji}^{n} + \sum_{k\neq j}\sum_{j\neq i} v_{kj} b_{Xji} p_{Xj}^{v} x_{ji}^{v} + \sum_{j\neq i} b_{Yji} p_{Yj} Y_{ji} \tag{6}$$

where p_{Xi}^{h} denotes the price charged by the representative HMNE with a plant in i. Let p_{Xj}^{n}, p_{Xj}^{v}, and p_{Yj} denote the prices charged by producers in j for goods X (NEs and VMNEs, respectively) and Y, respectively. In the second line of Eq. (6), the first three RHS terms denote factor income; the last three RHS terms denote tariff revenue redistributed lump-sum by the government in i back to the representative consumer. Let r_i denote the rental rate for physical capital in i, K_i is the physical capital stock in i (used at home or transferred abroad at a cost in units of capital of γ), w_{Si} (w_{Ui}) is the wage rate for skilled (unskilled) workers in i, and S_i (U_i) is the stock of skilled (unskilled) workers in i.

Maximizing (5) subject to (6) yields the domestic demand functions:

$$x_{ii}^{\ell} \geq (p_{Xi}^{\ell})^{\varepsilon-1} P_{Xi}^{-\varepsilon} \eta E_i; \quad \ell = \{n, h, v\} \tag{7}$$

where E_i is the income (and expenditure) of the representative consumer in country i from Eq. (6), and

$$P_{Xi} = \left[\sum_{j=1}^{3} n_j (t_{Xji} p_{Xj}^{n})^{\varepsilon} + \sum_{j=1}^{3} h_j (p_{Xi}^{h})^{\varepsilon} + \sum_{k\neq j}\sum_{j=1}^{3} v_{kj} (t_{Xji} p_{xj}^{v})^{\varepsilon}\right]^{\frac{1}{\varepsilon}} \tag{8}$$

is the corresponding CES price index. Following the literature, we assume all firms producing in the same country face the same technology and marginal costs and we assume complementary-slackness conditions (cf., Markusen, 2002). Hence, mill (or ex-manufacturer) prices of all varieties in a specific country are equal in equilibrium. Then, the relationship between differentiated goods produced in j and at home in i is:

$$\frac{x_{ji}}{x_{ii}} = \left(\frac{p_{Xj}}{p_{Xi}}\right)^{\varepsilon-1} t_{Xji}^{\varepsilon} (1 + b_{Xji})^{-1}. \tag{9}$$

Hence, from now on we can omit superscripts for both prices and quantities of differentiated products for the ease of presentation. It follows that homogeneous goods demand is:

$$\sum_{j=1}^{3} Y_{ji} \geq \frac{1-\eta}{p_{Yi}} E_i. \tag{10}$$

3.2. Differentiated goods producers

We assume that manufactures can be produced in all three countries using skilled labor, unskilled labor, and physical capital. Each country is assumed to be endowed with exogenous amounts of internationally immobile skilled and unskilled labor. Each country is also endowed with an exogenous amount of physical capital; however, physical capital can be transferred endogenously across countries by MNEs to maximize their profits, thus making endogenous determination of bilateral FDI flows, cf., Footnote 9. Differentiated goods producers operate in monopolistically competitive markets, similar to Markusen (2002, Ch. 6). Two assumptions used for our theoretical results that follow are the existence of a third, internationally-mobile factor – physical capital – *and* that any headquarters setup (fixed cost) requires home skilled labor – to represent R&D – and any plant setup requires the home country's physical capital — to represent the resources needed for a domestic or foreign direct investment.[13]

Assume the production of differentiated good X is given by a nested Cobb–Douglas-CES technology where F_{Xi} denotes production of these goods for both the domestic and foreign markets; we assume MNEs and NEs have access to the same technology. Let K_{Xi}, S_{Xi}, and U_{Xi} denote the quantities used of physical capital, skilled labor, and unskilled labor, respectively, in country i to produce X:

$$F_{Xi} = B(K_{Xi}^{\chi} + S_{Xi}^{\chi})^{\frac{\alpha}{\chi}} U_{Xi}^{1-\alpha}. \tag{11}$$

The specific form of the production function is motivated by two literatures. First, the Cobb–Douglas function provides a standard, tractable, and empirically relevant method of combining capital and labor; α denotes the share of "capital" in production. Second, early work by Griliches (1969) indicates that physical capital and human capital tend to be complements in production; recent evidence for this in the domestic literature is in Goldin and Katz (1998) and in the MNE literature is in Slaughter (2000). We nest a CES production function within the Cobb–Douglas function to allow for potential complementarity of physical and human capital; χ determines the degree of complementarity or substitutability.

NEs and MNEs differ in fixed costs. Each NE incurs only one firm (or headquarters) setup and one plant setup; each MNE incurs one firm setup (the cost of which is assumed larger than that of an NE, as in Markusen, 2002) and a plant setup for its home market and for each foreign market it enters endogenously. A horizontal MNE has headquarters at home and plants in three markets to serve them; it has no exports. A vertical MNE has headquarters at home and one plant abroad, which can export to any market. Maximizing profits subject to the above technology yields a set of conditional factor demands reported in a technical supplement (available from the authors).

3.3. Homogeneous goods producers

We assume homogeneous good (Y) is produced under constant returns to scale in perfectly competitive markets using only unskilled labor; assume the technology $Y_i = Ui$ (i=1,2,3). In the presence of positive trade costs, we assume country 1 is the numeraire; hence, $p_{Y1} = w_{U1} = 1$.

[13] Note it is not necessary that plants (firms) require only physical (human) capital to setup plants (firms); what is necessary is that setups of plants (firms) are *relatively* more physical (human) capital intensive, which we conjecture is true empirically. Also, the model is robust to assuming instead that plants (firms) require human (physical) capital for setups. The key is that two factors are used in setups, and that the two setups (firms and plants) have different relative factor intensities.

3.4. Profit functions, pricing equations, and the definition of FDI

Firms are assumed to maximize profits given the technologies and the demand relationships suggested above. The profit functions are:

$$\pi_i^n = (p_{Xi} - c_{Xi})\sum_{j=1}^{3} x_{ij} - a_{Sni}w_{Si} - a_{Kni}r_i \tag{12a}$$

$$\pi_i^h = \sum_{j=1}^{3}(p_{Xj} - c_{Xj})x_{jj} - a_{Shi}w_{Si} - a_{Khi}[3 + \sum_{j\neq i}\gamma_{ij}]r_i \tag{12b}$$

$$\pi_{ij}^v = (p_{Xj} - c_{Xj})\sum_{k=1}^{3} x_{jk} - a_{Svi}w_{Si} - a_{Kvi}[1 + \gamma_{ij}]r_i. \tag{12c}$$

Eq. (12a) is the profit function for each NE in i. Let c_{Xi} denote marginal production costs of differentiated good X in country i and the last two RHS terms represent, respectively, fixed human and physical capital costs for the NE producer. Eq. (12b) is the profit function for each HMNE in country i with three plants. The last two terms in (12b) represent fixed costs of each 3-country HMNE, a single fixed cost of home skilled labor to setup a firm and a fixed cost of home physical capital for each plant. Each foreign investment incurs an additional investment cost γ (say, policy or natural foreign direct investment barrier). Consequently, in the context of our model, the flow (and stock) of FDI of country i's representative three-country MNE in country j (if profitable) would be $a_{Khi}r_i(1+\gamma_{ij})$; in our model, international capital "mobility" is defined as country i's physical capital being used abroad (say, in j), but the factor rewards are earned in i, cf., Footnote 9. Eq. (12c) is the profit function for a vertical MNE with a headquarters in i and a plant in j; FDI from i to j is analogously $a_{Kvi}r_i$ $(1+\gamma_{ij})$.

A key element of our model is that – in each country – the numbers of NEs (type n), HMNEs (type h), and VMNEs (type v) are *endogenous* to the model. Two conditions characterize models in this class. First, profit maximization ensures markup pricing equations:

$$p_{Xi} \leq \frac{c_{Xi}(\varepsilon - 1)}{\varepsilon}. \tag{13}$$

Second, free entry and exit ensure:

$$a_{Sni}w_{Si} + a_{Kni}r_i \geq \frac{c_{Xi}(\varepsilon - 1)}{\varepsilon}\sum_{j=1}^{3} x_{ij} \tag{14a}$$

$$a_{Shi}w_{Si} + a_{Khi}[3 + \sum_{i\neq j}\gamma_{ij}]r_i \geq \sum_{j=1}^{3}\frac{c_{Xj}(\varepsilon - 1)}{\varepsilon}x_{jj} \tag{14b}$$

$$a_{Svi}w_{Si} + a_{Kvi}[1 + \gamma_{ij}]r_i \geq \frac{c_{Xj}(\varepsilon - 1)}{\varepsilon}\sum_{k=1}^{3} x_{jk}. \tag{14c}$$

3.5. Factor-endowment and current-account-balance constraints

We assume that, in equilibrium, all factors are fully employed and that every country maintains multilateral (though not bilateral) current-account-balance; endogenous bilateral current-account imbalances allow for endogenous bilateral FDI of physical capital. Following the established literature, this is a static model. The formal factor-endowment and multilateral current-account-balance constraints are provided in a technical supplement (available from the authors).

4. Calibration of the model

The complexity of the model (including the complementary-slackness conditions) introduces a high degree of nonlinearity, and it cannot be solved analytically. As in Markusen (2002), we provide numerical solutions to the model. As common, results may be sensitive to choice of parameters. Hence, we go to some effort to choose parameters and exogenous variables' values closely related to econometric evidence and empirical data. With three countries, we have several potential types of asymmetries, e.g., large vs. small GDPs, developed (DC) vs. developing (LDC) economies. To limit the scope, we focus here on bilateral flows between two developed economies, initially assuming the third economy (ROW) is also developed. However, in a robustness analysis later, we compare the results to those with a developing ROW. We use *GAMS* for our numerical analysis.

4.1. Values of (exogenous) factor-endowment, trade cost, and investment cost variables

We assume a world endowment of physical capital (K) of 240 units, skilled labor (S) of 90 units, and unskilled labor (U) of 100 units. Initially, each of the three-country's shares of the world endowments is 1/3; hence, all three countries have identical absolute and relative factor-endowments initially.

We appealed to actual trade data to choose initial values for transport costs (rather than choosing values arbitrarily, as typical). Using the bilateral final goods trade data from the United Nations' (UN's) COMTRADE data for weights, we calculated the mean final goods bilateral transport cost factors [(cif−fob)/fob] for intra-DC trade (7.6%) and trade between DCs and LDCs (20.2%).

We constructed initial values for tariff rates using Jon Haveman's TRAINS data for the 1990s. Tariff rates are available at the Harmonized System 8-digit level. Using the UN's COMTRADE data again, we classified tariff rates by final goods 5-digit SITC categories. We then weighted each country's 5-digit SITC final goods tariff to generate average tariffs at the country level for each year 1990–2000. Tariff rates were then weighted by aggregate bilateral final goods imports to obtain mean regional tariff rates, accounting for free trade agreements and customs unions. For final goods, the tariff rate was 1.1% for intra-DC trade and 9.7 (4.0) percent for trade from DCs to LDCs (LDCs to DCs).

Data on bilateral costs of investment and on policy barriers to FDI are not available. Carr et al. (2001) used a time-varying country "rating" score from the World Economic Forum's *World Competitiveness Report* that ranges from 0 to 100 (used in Section 7). However, *ad valorem* measures are not available across countries, much less over time. Consequently, we assumed values to represent informational costs and policy barriers to FDI between countries. We assumed initially a "tax-rate" equivalent (for γ) of 40% for intra-DC FDI and 90% for FDI flows between LDCs and DCs.

4.2. Utility and technology parameter values

Consider first the utility function. In Eq. (5), the only two parameters are the Cobb–Douglas share of income spent on differentiated products from various producers (η) and the CES parameter (ε) influencing the elasticity of substitution between differentiated products ($\sigma \equiv 1-\varepsilon$). Initially, we use 0.71 for the value of η, based upon an estimated share of manufactures trade in overall world trade averaged between 1990 and 2000 using 5-digit SITC data from the UNs' COMTRADE data set, which is a plausible estimate. The initial value of ε is set at -5, implying an elasticity of substitution of 6 among differentiated final goods, consistent with a wide range of recent empirical studies estimating the elasticity between 2 and 10, cf., Baier and Bergstrand (2001) and Head and Ries (2001a,b).

Consider next production function (11) for differentiated goods. Labor's share of differentiated goods gross output is assumed to be 0.8; the Cobb–Douglas formulation implies the elasticity of substitution between capital and labor is unity. As discussed earlier, Griliches (1969) found convincing econometric evidence that physical capital and skilled labor were relatively more complementary in production than physical capital and unskilled labor. Most evidence to date suggests that skills and physical capital are relatively complementary in production, cf., Goldin and Katz (1998) and Slaughter (2000). Initially, we assume $\chi=-0.25$, implying a technical rate (elasticity) of substitution of 0.8 [$=1/(1-\chi)$] and complementarity between physical and human capital.

As in Markusen (2002, ch. 5), a firm (or headquarters) setup uses only skilled labor. For national final goods producers, we assume a headquarters setup requires a unit of skilled labor per unit of output ($a_{Sn,1}=a_{Sn,2}=a_{Sn,3}=1$). As in Markusen, we assume "jointness" for MNEs; that is, services of knowledge-based assets are joint inputs into multiple plants. Markusen suggests that the ratio of fixed headquarters setup requirements for a horizontal or vertical MNE relative to a domestic firm ranges from 1 to 2 (for a 2-country model). We assume initially a ratio of 1.01 ($a_{Sh,1}=a_{Sh,2}=a_{Sh,3}=a_{Sv,1}=a_{Sv,2}=a_{Sv,3}=1.01$). Hence, to bias the theoretical results initially *in favor of multinational activity* (that is, in favor of MNEs completely displacing trade), we assume the additional firm setup cost of an MNE over a national firm is quite small. We assume that every plant (national or MNE) requires one unit of home physical capital ($a_K=1$); MNEs setting up plants abroad face additional fixed investment costs (γ), values for which were specified above.

5. Coexistence: the two-country, three-factor case

In this section, we show theoretically that bilateral intra-industry FAS and intra-industry trade can coexist when two countries have identical GDPs; HMNEs need not displace NEs completely. The third country is unnecessary to address coexistence, so we assume for now ROW has zero factor-endowments.

In the $2\times2\times2$ model of Markusen (2002) where the two countries have identical absolute as well as relative factor-endowments, the reason HMNEs displace final goods trade completely lies in "setups" of plants and firms *both* requiring human capital. The intuition is straightforward: if transport costs between two countries are sufficiently high and the ratio of headquarters scale economies (fixed costs) relative to plant scale economies (fixed costs) is large (small), then the relative cost for country *i* of supplying a foreign market *j* with goods from foreign affiliates of *i*'s HMNEs is low relative to exporting from *i*. While in some (asymmetric) combinations of the two countries' GDPs, HMNEs and NEs can coexist, a redistribution of the world's factor-endowments (with no relative factor-endowment change) *always* causes trade and FAS to move in opposite directions (due to scarcity of human capital).

Fig. 1a and b illustrate the issue for the $2 \times 2 \times 2$ case with no physical capital. First, we explain the figures' axes and labeling. Given the nonlinear, non-monotonic relationships in this class of GE models, we must rely on numerical solutions to the model to obtain "theoretical" relationships, using figures generated by the numerical model. In these figures, we focus on the bilateral relationships between two economies with identical relative factor-endowments and transaction cost levels. Fig. 1a and b illustrate the relationships between country economic size (referred to as the sum of GDPs of *i* and *j*) on the *y*-axis, similarity of economic size on the *x*-axis, and final good exports of NEs from *i* to *j* (the values of foreign affiliate sales of country *i*'s HMNEs in country *j*) on the *z*-axis in Fig. 1a (b).[14] The lines on the *y*-axis in the bottom plane range from 1 to 21. The *y*-axis indexes the joint absolute factor-endowments of countries *i* and *j*; line 1 denotes the smallest combination and line 21 denotes the largest combination. The *x*-axis is indexed from 0 to 1. Each line represents *i*'s share of both countries' absolute factor-endowments; the center line represents 50%, or *identical GDP shares for i and j*.[15]

These figures are three-dimensional analogues to a two-dimensional figure in Markusen (2002, p. 99, Fig. 5.5, bottom). Our figures allow a range of total economic sizes for the two countries, whereas Markusen's corresponding figure was for a (single) given total economic size for the two countries.[16] Suppose initially that country *j* has the smallest possible GDP (i.e., virtually no factor-endowments) and country *i* has virtually all of the two countries' endowments; this is the first point on the far RHS of Fig. 1a and b. When country *j* is very small, it is most profitably served by the only other country, *i*, by national firms exporting from *i*; hence, HMNEs based in *i* will not serve *j* with foreign affiliate sales ($FAS_{ij}=0$). Even a slightly larger GDP in country *j* (the point to the left of the initial one) is not enough to make human capital sufficiently scarce to warrant FAS. However, at higher levels (moving to the left), country *j* is a large enough market to warrant foreign affiliate production by country *i* given investment barriers, and exports by national firms in *i* are partially "displaced" by the scarcity of human capital. At equal GDPs, HMNEs displace trade *completely* ($FAS_{ij}>0$ and trade is zero).

The complete displacement of trade in the $2 \times 2 \times 2$ model was formalized in Markusen and Venables (2000, section 4). More accurately, in their model there exists only a (highly unlikely) *unique* combination of trade costs (t) and ratio of national-to-multinational-firm setup costs (f/g) where HMNEs and NEs can coexist when the two countries are symmetric and the equilibrium is in the factor-price-equalization set (where mill prices and marginal production costs are equal). In the two-factor case, only if f/g is chosen to equal precisely $(1+t^{1-\varepsilon})/2$ can NEs and HMNEs coexist, where ε is the elasticity of substitution between varieties in their model.[17]

[14] Gross bilateral FAS from *i* to *j* is for HMNEs in industry *X*. Exports from *i* to *j* denote the aggregate gross bilateral trade flow of good *X* and good *Y* products from *i* to *j*.

[15] The theoretical flows are generated using *GAMS* and the figures are generated using *MATLAB*. As with all such programs, *MATLAB* requires the *y* (*x*) axis to be indexed by the sum of exogenous absolute factor-endowments (*i*'s share of the two countries' absolute factor-endowments), not endogenous GDPs. As in Markusen (2002), we often use the terms GDPs, economic sizes, and absolute factor-endowments interchangeably; for brevity, the figures' axes are labeled using GDPs. Of course, in empirical work, instrumenting GDPs by absolute factor-endowments in gravity equations generates virtually identical GDP coefficient estimates, cf., Frankel (1997).

[16] Of course, in the two-country case, the sum of GDPs of *i* to *j* is world GDP. However, to maintain consistent labeling of axes with the three-country case later, we label the *y*-axis with sum of GDPs of *i* to *j*.

[17] Shim (2005) used this assumption to motivate exogenously coexistence of MNEs and NEs.

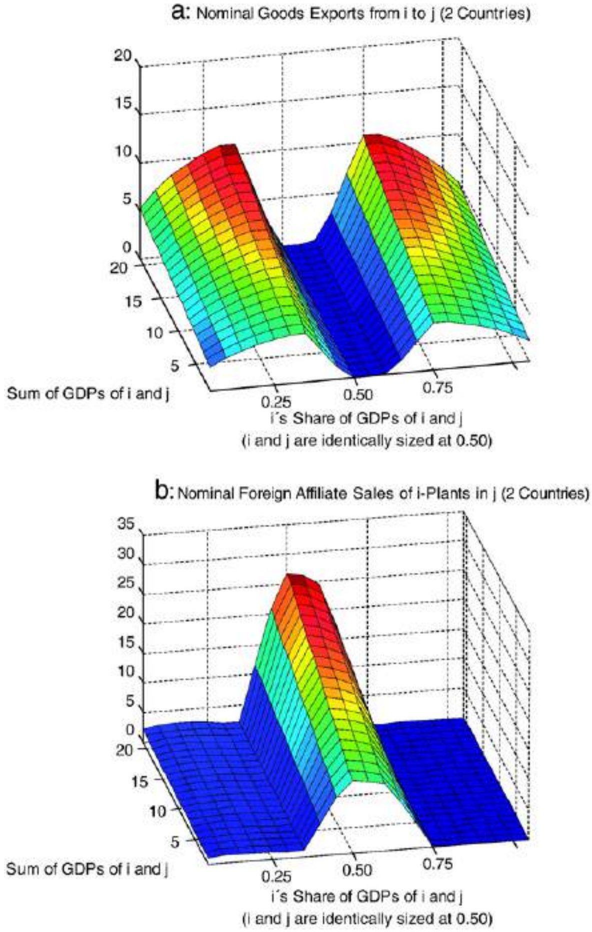

Fig. 1. a. Nominal goods exports from 1 to *j* (2 countries). b. Nominal foreign affiliate sales of *i*-plants in *j* (2 countries).

A resolution to this puzzle lies readily in introducing a third internationally-mobile factor and an assumption that home physical (human) capital is used to setup a plant (firm). The introduction of physical capital and the setup requirements can explain the observed *coexistence* of final goods trade (NEs) and FAS (HMNEs) when the two countries' GDPs are identical for a wide range of parameters. In simulations below, ROW has zero factor-endowments; conclusions result solely from introducing physical capital and the setup requirements.

Fig. 2a–c illustrate the implications for foreign affiliate sales, trade, and horizontal foreign direct investment, respectively, between a pair of developed-countries (*i*,*j*) with identical relative factor-endowments in our *three-factor*, two-country, two-good world. The most important point worth noting is, when the two countries have identical GDPs, NEs and HMNEs can *coexist*. However, similar to the 2 × 2 × 2 world, once HMNEs arise, FAS tends to *partially* displace trade and national exporters *no matter how large* are the GDPs of countries *i* and *j*, cf., middle columns in Fig. 2a and b. The intuition can be seen by starting at the far right side of Fig. 2b, when *j*'s GDP is small relative to *i*'s. When *j*'s GDP share becomes positive (via an exogenous reallocation of

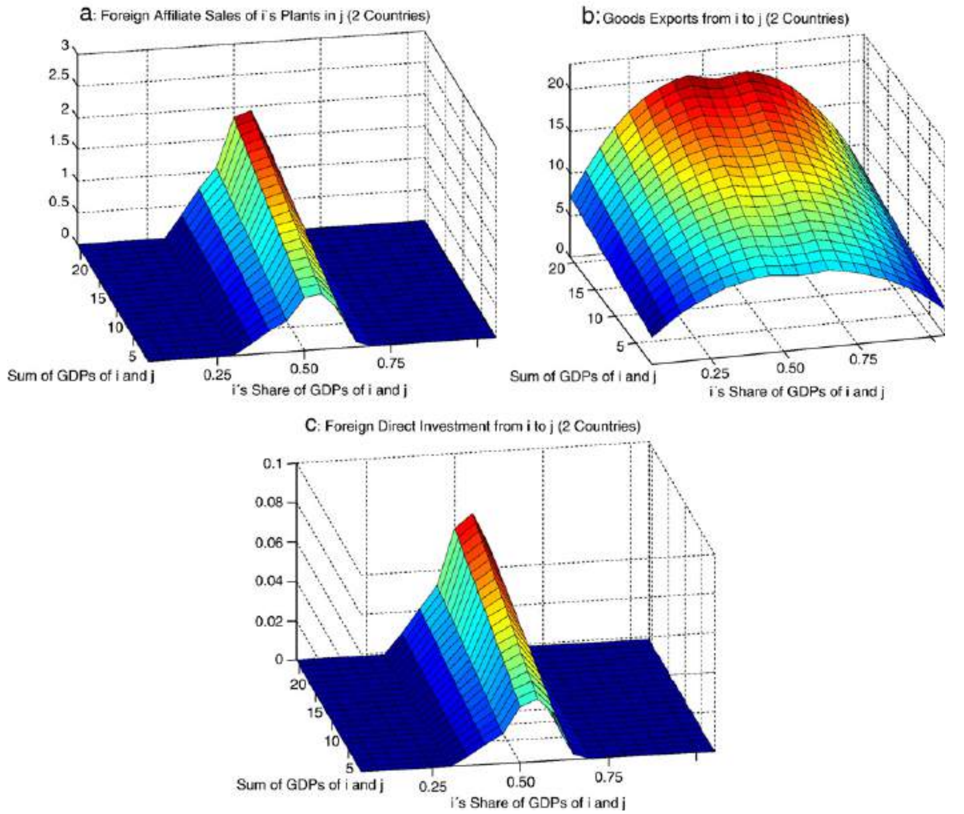

Fig. 2. a. Foreign affiliate sales of i's plants in j (2 countries). b. Goods exports from i to j (2 countries). c. Foreign direct investment from i to j (2 countries).

initial factor-endowments from i to j), j's relatively small market is most profitably served by exports from national firms in i and domestic production in j, identical to the $2 \times 2 \times 2$ case. Thus, i exports extensively to j.

As j's (i's) share of their combined GDPs gets larger (smaller), the number of exporters and varieties in i (n_i) contracts, cf., Fig. 3a. However, output per firm in j (x_{jj}, not shown), production, and overall demand expands proportionately more, such that exports from i to j increase, cf., Fig. 2b.

At some point (around 0.7 on the x-axis), as i's (j's) share gets even smaller (larger), it becomes more profitable for i to serve j's market using HMNEs (h_{2ij}) to avoid trade costs, cf., Fig. 3b. So FDI – physical capital – flows from i to j, so that both countries' markets can be more efficiently served by HMNEs based in i, cf., Fig. 2c. In the two-factor case with only human capital used to setup plants and firms, this would unambiguously increase the price of skilled (relative to unskilled) labor, displacing completely national firms in i. This need not occur here. While the replacement of NEs by MNEs requires more human capital for firm setups (as in the $2 \times 2 \times 2$ model), the net demand for human capital in i goes down because in our model plant setups require only physical capital. Since single-plant NEs are being increasingly replaced by multi-plant HMNEs as i and j become more similar in size (note the "dip" in the center columns of Fig. 3a when i's and j's shares are very similar), the relative demand for human capital falls and that for physical capital rises, lowering (raising) the price of human (physical) capital, cf., Fig. 3c and d.[18] This has two important implications for NEs not being *completely* displaced, and thus coexistence of NEs and HMNEs when i and j are identical. First, a higher price of physical capital in i (as i and j converge in size) raises the relative price of multi-plant HMNE firm setups, *reducing* the displacement of single-plant national firms (which serve markets via exports instead) and helping secure their coexistence with HMNEs. Second, a lower price of human capital in i lowers the price of HMNE and NE firm setups, also securing coexistence of both types of firms. As j's share approaches that of i, HMNEs continue to partially displace national firms. However, even with identical GDPs, HMNEs and national firms can *coexist* — a result precluded in the 2 × 2 × 2 model.

With physical capital, coexistence of HMNEs and NEs occurs over a wide range of economic sizes, GDP similarities, plant-to-firm setup costs, and trade costs, because of variable relative prices of human-to-physical capital, cf., Fig. 3c and d. We now demonstrate formally such coexistence for a wide parameter range. Consider the case of two symmetric economies where mill prices of the countries' differentiated products are identical ($p_{Xi}=p_{Xj}$), as in Markusen and Venables (2000, section 4). Begin with profit functions (12) from Section 3.4. With only two countries, there are no three-country HMNEs; we will have two-country HMNEs in equilibrium (h_2). With two symmetric economies in size and relative factor-endowments, there are no vertical MNEs (v). Hence, in equilibrium, zero profits ensure for the representative firm in i (and symmetrically in j):

$$\pi_n = (p_X - c_X)(x_{ii} + x_{ij}) - a_{Sn} w_S - a_{Kn} r = 0 \tag{15a}$$

$$\pi_{h2} = (p_{Xi} - c_{Xi})x_{ii} + (p_X - c_X)x_{jj} - a_{Sh2} w_S - a_{Kh2}[2 + \gamma]r = 0. \tag{15b}$$

Since mill prices are the same in i and j, we can use Eq. (9) to note:

$$\frac{x_{ij}}{x} = t_X^{1-\sigma}(1 + b_X)^{-1} \tag{16}$$

where, due to symmetry, $x=x_{ii}=x_{jj}$.

We can now solve for the necessary condition that ensures coexistence of NEs and HMNEs. Substituting Eq. (16) into Eq. (15a) and solving for $(p_X - c_X)x$ yields:

$$(p_X - c_X)x = (a_{Sn} w_S + a_{Kn} r)/[1 + t_X^{1-\sigma}(1 + b_X)^{-1}].$$

However, we know from Eq. (15b) that:

$$2(p_X - c_X)x = (a_{Sh2} w_S + a_{Kh2}[2 + \gamma]r).$$

Hence, after some algebra, the necessary condition for coexistence is:

$$[(a_{Sh2}(w_S/r) + a_{Kh2}(2 + \gamma)]/2 = [a_{Sn}(w_S/r) + a_{Kn}]/[1 + t_x^{1-\sigma}(1 + b_X)^{-1}].$$

Thus, unlike in Markusen and Venables (2000), NEs and HMNEs can coexist for two *identically-sized* economies across a wide variety of values of trade costs, investment costs, and plant-to-firm setup costs due to the endogeneity of the relative price of human-to-physical capital.[19]

[18] Recall, human and physical capitals are complements in *production* of differentiated goods, not in setups.

[19] We can also show that the addition of the third factor alone is not sufficient to ensure coexistence. Eliminating the physical capital requirement in plant setup costs (replacing it with human capital) leads to a unique solution result, as in Markusen and Venables (2000). However, it is not critical that physical (human) capital is used for plant (firm) setups; this assumption can be reversed, as long as the world endowment of skilled labor is increased (to ensure positive FDI and FAS).

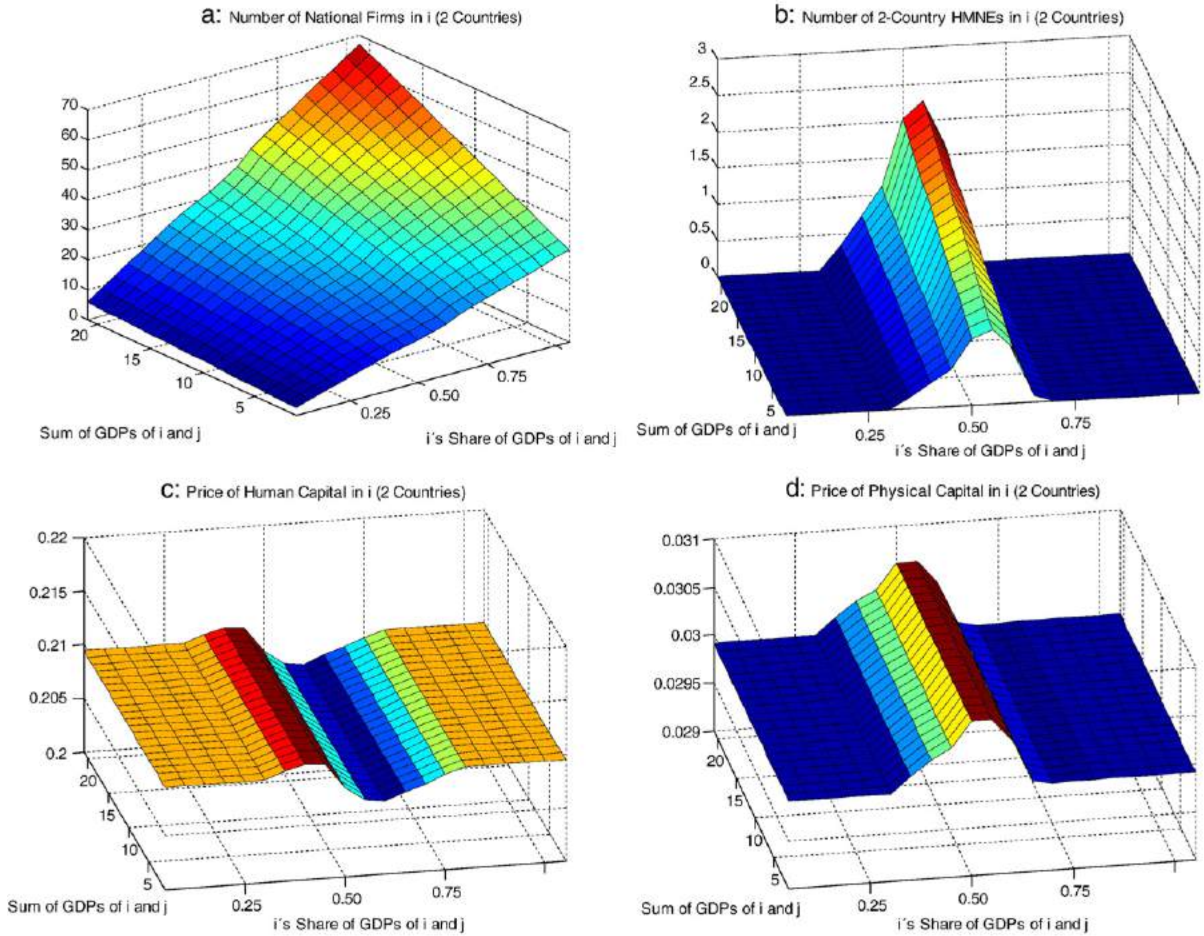

Fig. 3. a. Number of national firms in *i* (2 countries). b. Number of 2-country HMNEs in *i* (2 countries). c. Price of human capital in *i* (2 countries). d. Price of physical capital in *i* (2 countries).

As in this class of models, the results are, of course, sensitive quantitatively to levels of transport costs, investment costs, firm fixed costs, and plant fixed costs. However, these results hold for a wide range of values. Although the presence of physical capital and the assumption that plant setups use only physical capital generate the result that horizontal MNEs need not displace trade completely, the degree of substitutability or complementarity of bilateral trade and FDI with respect to economic size and similarity depends upon the presence of a "third country."

6. Complementarity: the three-country, three-factor case

While the gravity equation is familiar in algebraic form, it will be useful to *visualize* the frictionless "gravity" relationship. Fig. 4a illustrates the gravity–equation relationship between bilateral trade flows, GDP size, and GDP similarity summarized in Eq. (3) for an arbitrary hypothetical set of country GDPs ($N>2$). Of course, one notes a qualitative similarity of Fig. 4a to Fig. 2a–c.

However, there are (at least) three notable shortcomings that the *two-country* model cannot address. First, Fig. 2a–c represent bilateral *and* multilateral international flows; these figures depict a two-country world. With only two countries one cannot identify the effect of ROW on the bilateral flow from i to j; a theoretical foundation for the bilateral gravity equation needs at least three countries.

Second, with two countries the trade flow from i to j is *non-monotonic* in GDP similarity, cf., Fig. 2b. In a gravity equation, the trade flow is monotonic in economic similarity, cf., Fig. 4a. This conflicts with empirical results of trade and FDI gravity equations that imply bilateral trade and FDI are *complements* with respect to GDP size and similarity.

Third, in contrast to trade gravity equations, typical FDI gravity equation estimates find home country GDP elasticities *significantly larger* than host country GDP elasticities, cf., Table 1. This asymmetry cannot be explained using the two-country, three-factor model, cf., Fig. 2c.

In part 6.1, we address these three shortcomings. First, we explain why the introduction of a third country is necessary to generate "complementary" relationships between two countries' bilateral trade, FDI, and FAS flows with the two countries' GDP size and similarity. We address why the numbers of NEs and HMNEs also change in a complementary way with respect to changes in the two countries' economic size and similarity — even though NEs and HMNEs remain "overall" substitutes. Moreover, by introducing the third country, we can explain readily why typical empirical FDI gravity equations find home country (i) GDP elasticities significantly larger than host country (j) GDP elasticities. In part 6.2, we show the explicit relationship between size of the third country's (ROW's) GDP and trade, FDI and FAS of countries i and j, and suggest some "testable" hypotheses about the relationships between trade from i to j and FDI from i to j with GDP_{ROW}.

6.1. A theoretical rationale for estimating gravity equations of trade, FAS, and FDI

In this section, we introduce a third country – ROW – which initially is *identical* to the other two countries; the three countries are referred to as i, j, and ROW. Figs. 4b–d (5a–d) are the analogous "three-country" figures to two-country Figs. 2a–c (3a–d).[20] First, we note unsurprisingly in Fig. 4b–d that bilateral FAS, trade, and FDI from i to j, respectively (henceforth, FAS_{ij}, $Trade_{ij}$, and FDI_{ij}) are all positive monotonic functions of the *size* of the two countries' GDPs for any given share of i in i's and j's total GDP, as the gravity equation suggests (cf., Fig. 4a) for a given GDP_{ROW}. Note in Fig. 5a

[20] In the center cell of the x- and y-axes of Figs. 4 and 5, all three countries' GDPs are identical.

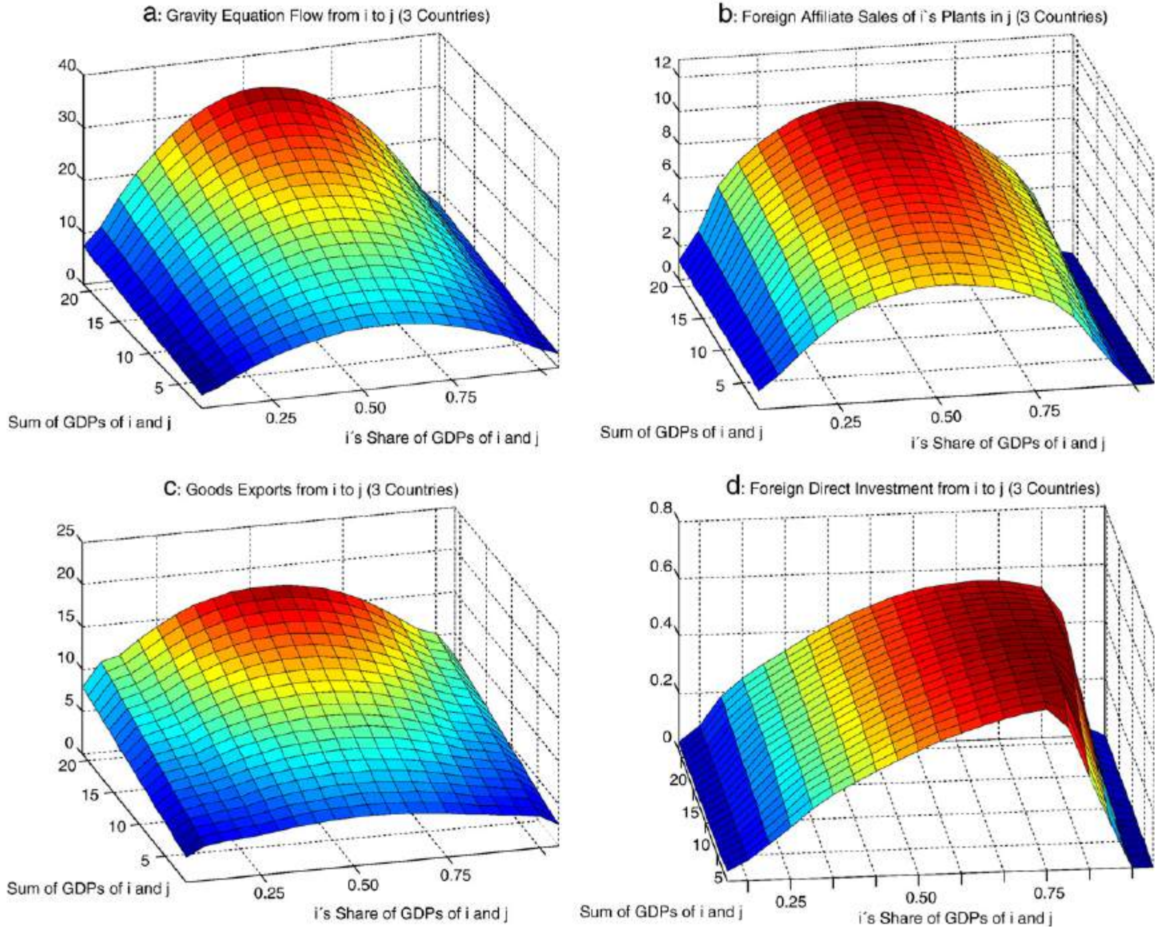

Fig. 4. a. Gravity equation flow from *i* to *j* (3 countries). b. Foreign affiliate sales of *i*'s plants in *j* (3 countries). c. Goods exports from *i* to *j* (3 countries). d. Foreign direct investment from *i* to *j* (3 countries).

and b that the numbers of both NEs and HMNEs increase with a larger joint economic size. More resources cover more setup costs of NEs and HMNEs; economies of scale and love of variety expand the number of total varieties produced for all three countries by both types of firms in *i*.

Second, and more interesting, even though NEs and HMNEs remain substitutes *overall*, the presence of the third country allows *both* bilateral FAS and trade from *i* to *j* to reach maximums as these two countries converge in GDP size, as a gravity equation would suggest; unlike the two-country case, FAS_{ij} and Trade_{ij} are "complements" with respect to similarity in *i*'s and *j*'s GDPs. Note first that Figs. 4b and c (5a and b) illustrate that FAS_{ij} and Trade_{ij} (NEs and HMNEs in *i*) remain substitutes overall. For any given size and similarity, the number of HMNEs (NEs) is larger (smaller) in the 3-country world than in the 2-country world; compare any cell in Fig. 5b (5a) with the corresponding cell in Fig. 3b (3a). Enlarging the world's economic size shifts the structure of all three economies toward relatively more HMNEs and fewer NEs; in this dimension, NEs and HMNEs are "overall" substitutes. With three countries, HMNEs surface more readily to avoid expensive trade costs; *i* can more profitably serve *j* and ROW by avoiding trade costs and investing in plants abroad. Note in Fig. 5b (4d) that *i*'s number of three-country HMNEs (*i*'s bilateral FDI from *i* to *j*) is positive even when *j*'s GDP is quite small relative to *i*'s. FAS_{ij} and FDI_{ij} (Trade_{ij}) is larger (smaller) in the 3-country relative to 2-country world. The overall substitution of FAS and FDI for trade also explains the "flattening" of the FAS (FDI) theoretical surface in Fig. 4b (d) relative to that in Fig. 2a (c); with more HMNEs in the 3-country world, FAS spreads out over a wider range of GDPs sizes and similarities since FDI occurs even when *j*'s GDP is small.

Yet, if the presence of a third country causes FAS_{ij} to tend to substitute more for Trade_{ij} for any given level of GDP size and similarity, why do both increase as the country pair's GDPs become more identical with three countries but not with two? As the GDPs of countries *i* and *j* become more identical, Fig. 5c and d illustrate that the price of human (physical) capital tends to fall (rise). However, the larger percentage fall (rise) in the price of human (physical) capital with three countries than with two dampens the increase in HMNEs and the decrease in NEs more with three countries.[21] The dampening of the increase in the number of HMNEs contributes to "flattening" the FAS and FDI surfaces with three countries compared with two. The dampening of the decrease in the number of NEs is sufficient to cause Trade_{ij} to be *monotonic* in size and similarity of the GDPs of countries *i* and *j* in the three-country case. The combined effect is to cause FAS_{ij} *and* Trade_{ij} to – on net – covary positively and monotonically with respect to country *i*'s and *j*'s GDP size and similarity in the three-country world. This is the main economic reason behind the gravity equation working for FAS, FDI, and trade flows *simultaneously*!

Consequently, *i*'s FAS in *j and* trade from *i* to *j* are maximized when these two countries have identical GDPs (Fig. 4b and c), as a result of a third country and third factor. Thus, the theoretical relationships between bilateral trade, FDI, and FAS flows and GDPs in Fig. 4b–d are "well approximated" by the gravity relationship. However, note that the theoretical surface for FDI in Fig. 4d is distinctly asymmetric relative to the other surfaces. FDI from *i* to *j* is maximized when *i*'s share of the two countries' GDPs is larger than *j*'s share. It can be shown mathematically that such a surface implies the home country GDP elasticity of FDI *has to be larger* than the host country GDP elasticity. Thus, with three countries one should expect the empirical result found in specification (3) in Table 1, that is, a larger home country (*i*) GDP elasticity than host country (*j*) GDP elasticity for FDI_{ij}.

[21] For instance, as *j*'s share of *i*'s and *j*'s GDPs goes from 0 to 1/2 (for the smallest sum of GDPs of *i* and *j*), the fall in the price of *i*'s human capital is 7.8 (1.2) percent for the three-(two-) country case and the rise in the price of *i*'s physical capital is 16.7 (2.0) percent for the three-(two-) country case.

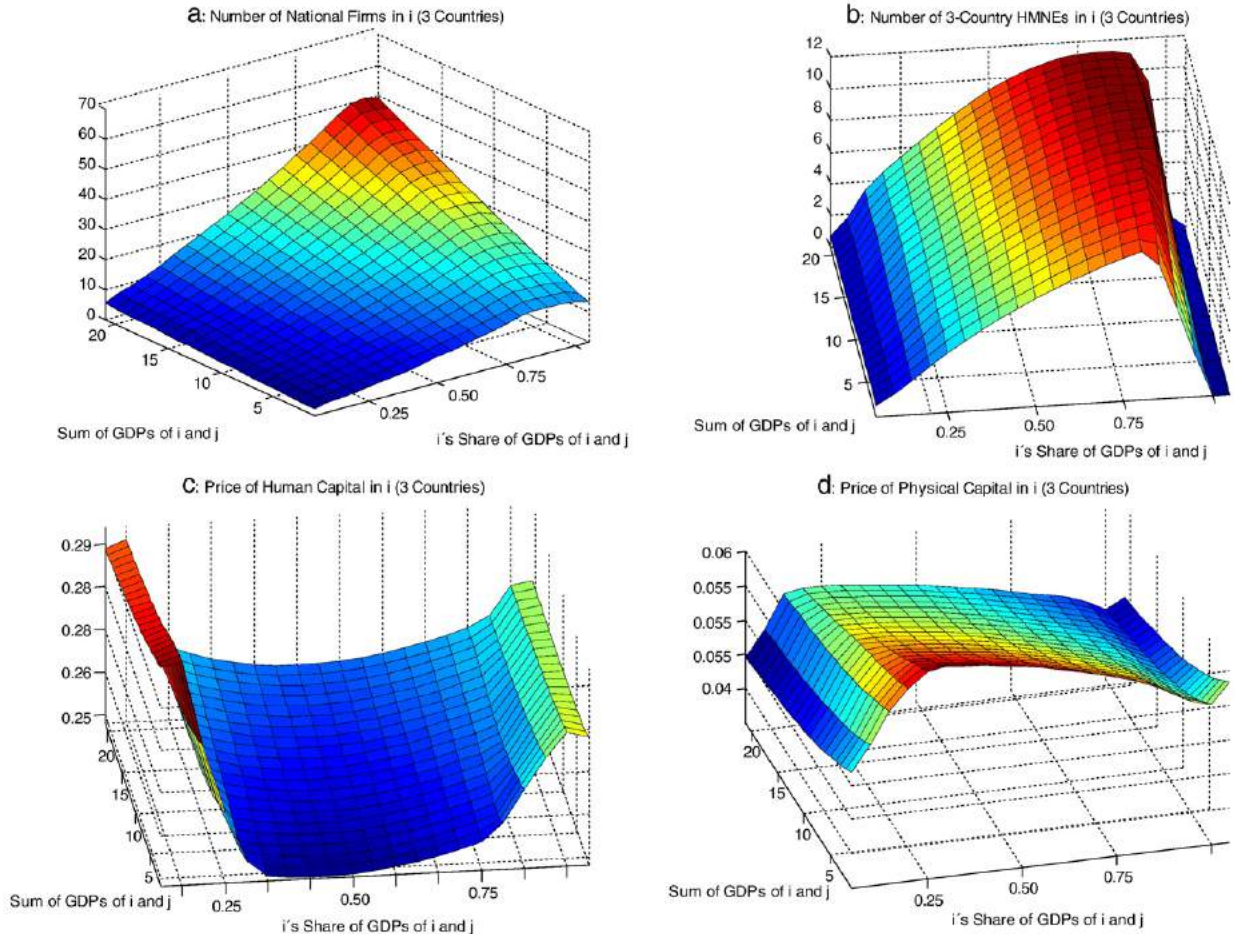

Fig. 5. a. Number of national firms in *i* (3 countries). b. Number of 3-country HMNEs in *i* (3 countries). c. Price of human capital in *i* (3 countries). d. Price of physical capital in *i* (3 countries).

However, "eyeballing" surfaces 4c and 4d *cannot* explain the empirical results using (the more econometrically appropriate) fixed-effects results in specifications (5)–(8) in Table 1. In Section 7 we move beyond "ocular-metrics" to provide a quantitative theoretical rationale for GDP elasticities estimated using fixed effects.

6.2. The effect of ROW on bilateral trade

In this part, we motivate potentially "testable" propositions about the empirical correlations between FDI_{ij} and GDP_{ROW} and between $Trade_{ij}$ and GDP_{ROW}. Fig. 6a–c show the relationships between the ROW's GDP (x-axis), the combined GDPs of i and j (y-axis), and $Trade_{ij}$, FAS_{ij}, and FDI_{ij}, respectively (z-axis). Here, the GDPs of countries i and j are set to be identical to each other, but their combined GDPs can vary. The cell in the middle row of the first column is defined to be identical absolute (and relative) factor-endowments for countries i, j and ROW. Suppose all three countries have identical absolute factor-endowments; when countries i and j have two-thirds of the world's GDP (for which there is no empirical counterpart, as will be shown), an increase in ROW's GDP makes FDI_{ij} relatively more attractive and $Trade_{ij}$ less attractive. As ROW's size initially increases, the profitability in country i from having more three-country HMNEs rises to avoid expensive transport costs to either j or ROW; hence, FDI from i to j (as well as i to ROW)

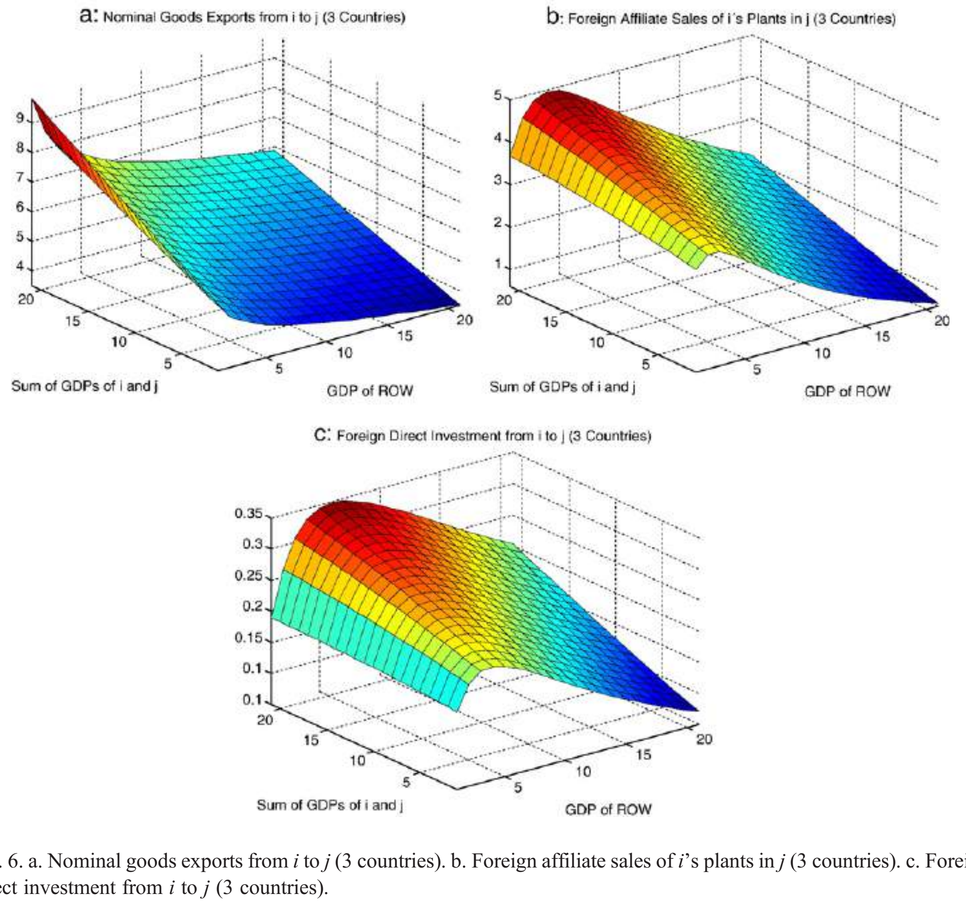

Fig. 6. a. Nominal goods exports from i to j (3 countries). b. Foreign affiliate sales of i's plants in j (3 countries). c. Foreign direct investment from i to j (3 countries).

increases, substituting for exports from i to j (Trade_{ij} falls). However, as ROW continues to increase in economic size (moving to the right along the middle row of the x–y plane), the relative market sizes of i and j shrink, attracting less exports from i to j by NEs and countries i and j become less attractive markets in size to be served efficiently by HMNEs in i and j.

The model suggests two potentially "testable" hypotheses. First, Trade_{ij} should decrease monotonically as ROW's GDP increases, consistent with previous formal theoretical models for the trade gravity equation ignoring MNEs and FDI. Second, the model suggests a *quadratic* relationship between FDI_{ij} and GDP_{ROW}; this hypothesis is novel. Clearly, a positive or negative relationship depends on relative GDPs of i, j, and ROW.

However, examination of actual GDPs suggests the likely relationship to dominate empirically for the second hypothesis is negative. Recall first that the x-axis has 21 columns. In Fig. 6c, GDP_{ROW} ranges from the same size as i (or j) in column 1 to 10 times GDP_i (or GDP_j) in column 21. Thus, column 2 represents $\text{GDP}_{\text{ROW}}=1.5*\text{GDP}_i$ and column 3 represents $\text{GDP}_{\text{ROW}}=2*\text{GDP}_i$, or $\text{GDP}_{\text{ROW}}=\text{GDP}_i+\text{GDP}_j$. Thus, the model hypothesizes that FDI_{ij} and GDP_{ROW} should be negatively (positively) correlated for observations where $\text{GDP}_i+\text{GDP}_j<\text{GDP}_{\text{ROW}}$ ($\text{GDP}_i+\text{GDP}_j>\text{GDP}_{\text{ROW}}$).

Does GDP data provide any guidance on the relevant hypothesis? For the last year of our sample (2000), the combined nominal GDP of the two largest economies in the world – the United States and Japan – was US$14.52 trillion; this was less than *half* of world GDP ($31.43 trillion, using the 160 largest economies) and hence less than GDP_{ROW} (US$16.91 trillion). Consequently, *all* empirical observations lie in the negatively-sloped region of Fig. 6c. Thus, the model implies two "testable" propositions: Trade_{ij} and FDI_{ij} should *both* be negatively related to ROW GDP. We evaluate these later empirically.

7. A quantitative description of the theoretical gravity equations of trade and FDI

The diversity of gravity-like relationships in specifications (5)–(8) in Table 1 suggests a need for a more quantitative "description" of the theoretical relationships in Fig. 4c and d.[22] To move beyond "ocular-metrics," in part 7.1 we describe a new methodology to describe *quantitatively* the relationship between each of the "theoretical flows" and the exogenous RHS variables, similar to how a regression describes empirical relationships. In part 7.2, we present the results of these exercises for describing the theoretical relationships between GDP size, GDP similarity, and the various flows, and compare these to the empirical results in Table 1. In part 7.3, we use the numerical GE model to illustrate theoretical relationships between trade costs, investment costs, RTAs, trade flows, and FDI. In part 7.4, we use data on ROW GDP and bilateral trade and investment costs to extend empirical results to determine if the "testable" propositions regarding ROW GDP, trade costs, investment costs, and trade and FDI flows hold. In part 7.5, we discuss some sensitivity analyses that were conducted.

7.1. Methodology

The key notion behind the methodology that follows is to describe quantitatively using a single regression equation a nonlinear theoretical surface relating bilateral flows to GDP size and

[22] The diversity of quantitative gravity relationships reflects the model's "unconditional" GE structure, i.e., non-separability of production, consumption, trade and FDI allocations. This differs from a standard gravity equation, such as Eq. (1), which is consistent with the "conditional" GE structure emphasized in Anderson and van Wincoop (2004, pp. 706–707), which employs a critical "trade-separability" assumption, resulting in unit GDP elasticities. See also Footnote 4.

similarity generated by a numerical GE model. We have empirical data on GDP sizes and similarities. Our surfaces map certain "theoretical" levels of bilateral (trade, FDI, or FAS) flows to various GDP sizes and similarities. However, since it is a nonlinear surface, there is an "approximation error" relating the level of the flow to GDP sizes and similarities. We treat this error term as normally distributed (and can confirm that assumption) so that a regression of the "theoretical" flow on GDP sizes and similarities describes the surface quantitatively.

Our methodology using regressions on "theoretical" flows is described most transparently using several steps. First, consider the bottom plane in Fig. 4c. The x-axis indexes country i's share of i's and j's total GDP, ranging from 0 to 1. The y-axis represents the sum of GDPs of countries i and j (GDPT).[23] The x-(y-) axis actually consists of 99 columns (600 rows); however, if we had used 99 columns (600 rows) in Fig. 4c, the entire figure would be black due to too many lines on the grid. For readability, the figures shown aggregate the 99 columns (600 rows) into 21 columns and rows so each cell's color content is visible. On the y-axis, the initial level of the country pair's GDP sum is set to one. The next line's GDP sum is six percent larger than the first line's GDP, the third line's GDP sum is six percent larger than the second line's GDP, and so on for 600 entries. Thus, GDP sums vary from 1 to (approximately) 37. This generates 59,400 ($=99\times600$) cells with "theoretical" observations on (i's and j's) GDP sums, GDP shares, and the level of trade flow, FDI, or FAS implied by the model.

We could run a regression of the numerical GE model's estimates of the (log) levels of the theoretical trade flows from i to j on (logs of) i's and j's GDP sum and similarity (E_i+E_j and $s_i s_j$, respectively) using each "cell" once as an observation, similar to Fig. 4a. The "error terms" would be the approximation (or rounding) errors generated from the nonlinearity of the surface and behave as if randomly distributed.[24] However, running such a regression would have limited economic meaning since each cell has the same weight in the distribution. Some of the cells (in terms of GDP totals and similarity) might not even be observed empirically!

Consequently, we create a more relevant distribution for the cells by benchmarking the (bottom) plane to empirical data on GDPs from *World Development Indicators*. We then fill each cell based upon observations of GDP sums and i's share of their GDP for that cell; that is, we "match" theoretical values of GDP size and similarity to their empirical observations for each country pair. Using our sample of the 17 most developed OECD countries' annual GDPs for 1990–2000, we fill the cells with the observations corresponding to each country pair with a certain GDP sum and GDP share (i's share of the GDP of i and j). For example, along the x-axis the first column includes all observations in the panel for which i's GDP share is between 0 and (approximately) 1% of a country pair's total GDP (recall, there are actually 99 columns). Along the y-axis, the first row (of actually 600 rows) includes all observations in the panel for which the GDP sum is smallest and up to 6% higher than that. The largest GDP sum in the sample (i.e., U.S.–Japan GDP sum for year 2000) amounts to about 37 times that of the smallest country pair. Thus, we have created a distribution of frequencies of cells based upon GDP sums and similarities using cross-section and time-series data. We then estimate the regression relationship between the *theoretical* value for a country pair's flow from i to j (using separately each of the three flows) based upon the model with the values for the GDP sums and i's GDP shares based on actual GDP data.[25]

[23] GDPs are, of course, an endogenous variable. The exogenous variables are the absolute endowments of unskilled labor, skilled labor, and physical capital. Simulations using these variables yield similar outcomes.

[24] We could have even added a stochastic error term explicitly; this would not have changed our results.

[25] Not every one of the 59,400 theoretical cells will be used. Also, we exclude all zero theoretical flows. The only difference between our regressions and typical regression analysis using empirical data on both sides of the regression is that the LHS variable is from the theoretical surface; RHS variables are random variables and the error term is normally distributed. Note, we are *not testing* the theory; we are simply describing the theory quantitatively.

7.2. Regression results using theoretical flows

Specifications (9)–(12) in Table 1 are the results of running the regressions using "theoretical" flows from i to j. Specification (9) reveals that theoretical trade flows from i to j are positively related to the countries' GDPs. The results are qualitatively identical to those in fixed-effects specification (5) in Table 1 using *actual* trade flows. The exporter GDP elasticity is smaller than the importer GDP elasticity in both specifications. A comparison of specification (10) using theoretical flows with fixed-effects specification (6) using actual trade flows shows that the results are qualitatively similar for GDP size and similarity. GDP similarity elasticities are nearly the same.

Specifications (11) and (12) in Table 1 report the coefficient estimates using the analogous specifications for "theoretical" FDI flows. Notably, the home country GDP elasticity is larger than the host country GDP elasticity in specification (11), consistent with fixed-effects specification (7) using actual FDI flows. Moreover, the GDP similarity elasticity in specification (12) is comparable to that in specification (8) using actual FDI flows. However, the GDP size elasticities are much larger in specifications (6) and (8) using empirical trade and FDI flows than in corresponding specifications (10) and (12) using theoretical flows. The relatively high GDP size elasticities may be sensitive to omitted variables or the particular period used for empirics (1990–2000); our time series was constrained by bilateral trade and investment cost data availability. We address potential omitted variables bias in part 7.4.

7.3. Effects of trade and investment costs, regional trade agreements, and relative factor-endowments

Of course, one of the most important variables in any trade or FDI gravity equation is bilateral distance, typically included as a proxy for trade or investment costs, respectively. Both trade and investment costs are included explicitly in our theoretical model (as in Markusen, 2002). Fig. 7a (b) shows the relationship between bilateral trade (FDI) flows, trade costs, and investments costs in our model. As expected, Trade_{ij} (FDI_{ij}) is negatively related to trade (investment) costs between i and j. Moreover, the model suggests, as common in this literature, that Trade_{ij} (FDI_{ij}) is *positively* related to investment (trade) costs between i and j. These "testable" propositions are evaluated empirically later.[26]

Fig. 7c and d illustrate the effect of introducing a bilateral regional free trade agreement (RTA) between countries i and j in our three-country model. The theory suggests that a reduction in bilateral trade costs is associated with a higher level of bilateral trade, cf., Fig. 7c. However, a reduction in trade costs makes bilateral investment less economical, and should tend to reduce bilateral FDI, cf., McCulloch (1985, 1993). Fig. 7d confirms that the model suggests that a free *trade* agreement between i and j should be associated with a lower level of bilateral investment from i to j. (Fig. 7c–d also provide a theoretical rationale for the multiplicative relationships between measures of trade and investment costs, GDPs, and trade and FDI flows in typical gravity equations.) In Section 7.4, we evaluate empirically whether trade (investment) is positively (negatively) related to RTA.

[26] The relationships shown in Fig. 7a and b are for the case of three countries that are identical in absolute and relative factor-endowments, as in Markusen (2002, sections 5.5). If relative factor-endowments were allowed to differ, such as in Markusen (2002, sections 8.5 and 8.6), we would obtain non-monotonic relationships. However, our purpose in Fig. 7a and b is to motivate expected empirical relationships for trade and FDI flows with trade and investment costs in the context of an empirical gravity equation where the interaction of such costs and relative factor-endowments is assumed to be minimized by the logarithmic transformation.

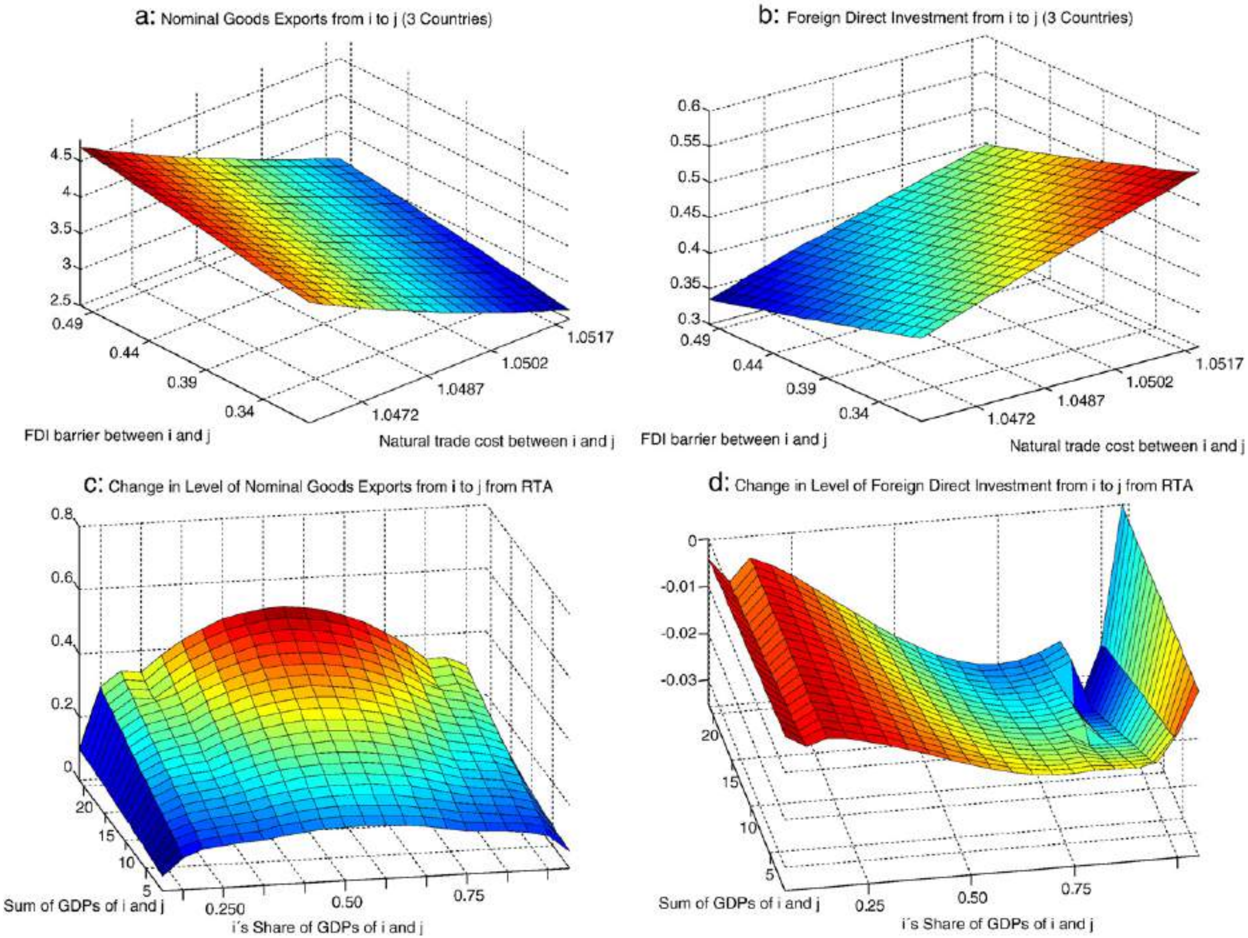

Fig. 7. a. Nominal goods exports from *i* to *j* (3 countries). b. Foreign direct investment from *i* to *j* (3 countries). c. Change in level of nominal goods exports from *i* to *j* from RTA. d. Change in level of foreign direct investment from *i* to *j* from RTA.

Up to now, countries have had identical relative factor-endowments. In the interest of robustness, we allowed ROW to be a developing country by having a larger (smaller) initial share of the world's unskilled labor (human and physical capital). With this change, vertical MNEs surface. However, the results are largely robust to this change. The surfaces examined in Figs. 4–7 are qualitatively (not quantitatively) identical with a developed or developing ROW; figures are omitted due to space constraints.

Also, our theoretical surfaces assume countries *i* and *j* have identical relative factor-endowments. As discussed in Carr et al. (2001), Blonigen et al. (2003), and Braconier et al. (2005), theoretical relationships between relative factor-endowments of countries *i* and *j* and trade, FAS, and FDI flows from *i* to *j* are interesting as well. However, examining those theoretical relationships for our three-factor model is well beyond the scope of this paper and the subject of future research. Nevertheless, even developed-country pairs are likely to have different relative factor-endowments. To ensure the empirical results that follow account for variation in relative factor-endowments between *i* and *j*, we include relative factor-endowments of country pairs.

7.4. Final empirical results

The previous sections suggest three potentially "testable" propositions that can be investigated empirically using time-series (within) variation. First, Section 6.2 suggested Trade_{ijt} should be negatively related to GDP_{ROW} and – based upon the actual size distribution of the world's GDPs – FDI_{ijt} should also be related negatively to GDP_{ROW}. Second, Section 7.3 suggested Trade_{ijt} (FDI_{ijt}) should be negatively related to bilateral trade (investment) costs and positively related to bilateral investment (trade) costs. Third, Section 6.3 suggested Trade_{ijt} (FDI_{ijt}) should be positively (negatively) related to an RTA dummy.

Specifications (13)–(16) in Table 1 report the results using bilateral fixed effects of estimating the same specifications similar to (5)–(8) but now adding ROW GDP, bilateral trade (cif−fob) and investment costs, and relative factor-endowments of countries *i* and *j*. Trade and investment costs data were described in Section 4; factor-endowment data are from Baier et al. (2004). First, we find that Trade_{ij} and FDI_{ij} are both statistically significantly negatively related to GDP_{ROW} as expected. Second, we find that Trade_{ij} (FDI_{ij}) is negatively related to bilateral trade (investment) costs as expected and the coefficient estimate is statistically significant at the 10% level (two-tailed t-tests). The relationship between Trade_{ij} (FDI_{ij}) and bilateral investment (trade) costs is not correctly signed, but these cross-price elasticities are much smaller than the own-price elasticities. Third, due to our data set (which was limited by investment cost data), we have no time-series variation in the free trade agreement dummy variable. However, we do have cross-section variation. A reexamination of Table 1's empirical gravity equation coefficient estimates for the dummy variable representing a free trade agreement (=1 if RTA present, =0 otherwise) using panel data in specifications (1)–(4) shows that a reduction in "trade costs" (in the form of the presence of a free trade agreement) is correlated with a higher level of bilateral trade – but a lower level of bilateral investment – between a pair of countries, consistent with our theory. Finally, while theoretical predictions for the relationships between relative factor-endowments of country *i* relative to *j* on flows are beyond the scope of this paper, we find that FDI_{ij} is positively and statistically significantly related to the (log) physical–capital-to-unskilled-labor (K/U) ratios (in *i* relative to *j*) and the (log) human-capital-to-unskilled-labor (S/U) ratios. Trade_{ij} is negatively related to the relative K/U ratios, but is positively and statistically significantly related to the relative S/U ratios.

7.5. Sensitivity of results to parameter selections and exogenous variables

Of course, the potential number of alternative calibrations of our numerical GE model is virtually unlimited. Paper space constraints prevent an exhaustive analysis of permutations from our base calibration. Nevertheless, we evaluated the sensitivity of the model's theoretical surfaces to assuming setups of headquarters (plants) require human (physical) capital and to varying trade-relative-to-investment costs, MNE/NE ratio firm setup costs, the capital–skill elasticity of technical substitution, and the elasticity of substitution of differentiated goods in utility. A complete discussion of these findings is in a technical supplement (at http://www.nd.edu/~jbergstr).

8. Conclusions and directions for future research

The introduction of a third internationally-mobile factor to the standard $2 \times 2 \times 2$ model allows us to resolve fairly readily a puzzle in the modern knowledge–capital literature on MNEs (without heterogeneous productivities among firms) that foreign affiliate sales among two identical economies completely displace their intra-industry trade. The introduction of a third country suggests a formal *N*-country theoretical rationale for estimating gravity equations of FDI flows and foreign affiliate sales, in a manner consistent with estimating trade gravity equations.

Our future research agenda is quite large. We intend to extend our methodological framework to examine theoretically the economic determinants of trade, FDI, and FAS flows *between* developed and developing countries, and *among* developing countries. We plan to incorporate intermediate goods to better understand the role of "outsourcing" for NEs and MNEs. We intend to investigate further the role of physical capital for better understanding the empirical relationships between relative factor-endowments and the pattern of MNE behavior as raised in Carr et al. (2001) and Blonigen et al. (2003).

Acknowledgements

We deeply appreciate extensive and insightful comments on earlier drafts by Eric van Wincoop, Bruce Blonigen, Ron Davies, Kishore Gawande, Amy Glass, David Hummels, Keith Maskus, Jim Markusen, Howard Schatz, Deborah Swenson, Chong Xiang, two anonymous referees, and participants at presentations at Purdue University, Syracuse University, Texas A&M University, the European Economic Association meetings (2005), and the Empirical Investigations in International Trade conference at UC-Santa Cruz (2004). Bergstrand acknowledges research funding from the Mendoza College of Business and Kellogg Institute for International Studies at Notre Dame and Egger acknowledges research funding from the Austrian Science Fund through grant P17713-G05.

Appendix A. Data appendix

Bilateral trade flow data are from the UN World Trade Database for 1990–2000. Bilateral (outbound) FDI stock data are from UNCTAD (country profiles) for same years. GDPs are from *World Development Indicators* (2004). Bilateral distance was computed using "great circle" distances. Dummy variables for adjacency and common language are from CEPII (Centre d'Etudes Prospectives et d'Informations Internationales). The 17 OECD countries include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and United States.

References

Anderson, James E., van Wincoop, Eric, 2003. Gravity with gravitas: a solution to the border puzzle. American Economic Review 93 (1), 170–192 (March).

Anderson, James E., van Wincoop, Eric, 2004. Trade costs. Journal of Economic Literature 42 (3), 691–751 (September).

Baier, Scott L., Bergstrand, Jeffrey H., 2001. The growth of world trade: tariffs, transport costs, and income similarity. Journal of International Economics 53 (1), 1–27 (February).

Baier, Scott L., Bergstrand, Jeffrey H., 2007. Do free trade agreements actually increase members' international trade? Journal of International Economics 71 (1), 72–95 (February).

Baier, Scott L., Dwyer, Gerald, Tamura, Robert, 2004. How important are capital and total factor productivity for economic growth? Economic Inquiry 44 (1), 23–49 (January).

Baldwin, Richard E., Ottaviano, Gianmarco, 2001. Multiproduct multinationals and reciprocal FDI dumping. Journal of International Economics 54, 429–448.

Barba Navaretti, G., Venables, Anthony J., 2004. Multinational Firms in the World Economy. Princeton University Press, Princeton, NJ.

Bergstrand, Jeffrey H., Egger, Peter, 2004. A theoretical and empirical model of international trade and foreign direct investment with outsourcing: part I, developed countries. Working Paper. April.

Bergstrand, Jeffrey H., Egger, Peter, 2006. A knowledge-and-physical–capital model of international trade, foreign direct investment, and foreign affiliate sales: developed countries. Working Paper. Jan.

Blonigen, Bruce A., 2001. In search of substitution between foreign production and exports. Journal of International Economics 53 (1), 81–104 (February).

Blonigen, Bruce A., 2005. A review of the empirical literature on FDI determinants. Atlantic Economic Journal 33 (4), 383–403 (December).

Blonigen, Bruce A., Davies, Ronald B., Head, Keith, 2003. Estimating the knowledge–capital model of the multinational enterprise: comment. American Economic Review 93 (3), 980–994.

Blonigen, Bruce A., Davies, Ronald B.,Waddell, Glen R., Naughton, H., in press. FDI in space: spatial autoregressive relationships in foreign direct investment. European Economic Review.

Borga, Maria, Yorgason, Daniel R., 2002. Direct investment positions for 2001: country and industry detail. Survey of Current Business. U.S. Department of Commerce, pp. 25–35. July.

Braconier, Henrik, Norback, Pehr-Johan, Urban, Dieter, 2005. Reconciling the evidence on the knowledge–capital model. Review of International Economics 13 (4), 770–786 (September).

Carr, David, Markusen, James R., Maskus, Keith E., 2001. Estimating the knowledge–capital model of the multinational enterprise. American Economic Review 91 (3), 693–708 (June).

Egger, Peter, 2000. A note on the proper econometric specification of the gravity equation. Economics Letters 66 (1), 25–31.

Ekholm, Karolina, Forslid, Rikard, Markusen, James R., in press. Export-platform foreign direct investment. Journal of the European Economic Association.

Feenstra, Robert C., 2004. Advanced International Trade: Theory and Evidence. Princeton U. Press, Princeton, NJ.

Frankel, Jeffrey A., 1997. Regional Trading Blocs in the World Economic System. Institute for International Economics, Washington, DC.

Goldin, Claudia, Katz, Lawrence F., 1998. The origins of technology-skill complementarity. Quarterly Journal of Economics 113 (3), 693–732.

Grazalian, Pascal L., Furtan, W. Hartley, 2005. The border effects: recognizing the role of FDI. Working Paper. University of Saskatchewan.

Griliches, Z., 1969. Capital–skill complementarity. Review of Economics and Statistics 51 (4), 465–468 (Nov.).

Head, Keith, Ries, John, 2001a. Overseas investment and firm exports. Review of International Economics 9 (1), 108–122 (February).

Head, Keith, Ries, John, 2001b. Increasing returns versus national product differentiation as an explanation for the pattern of U.S–Canada Trade. American Economic Review 91 (4), 858–876 (September).

Helpman, Elhanan, 2006. Trade, FDI, and the organization of firms. Journal of Economic Literature 44 (3), 589–630 (September).

Helpman, Elhanan, Krugman, Paul, 1985. Market Structure and Foreign Trade. MIT Press, Cambridge, MA.

Helpman, Elhanan, Melitz, Marc, Yeaple, Stephen, 2004. Export versus FDI with heterogeneous firms. American Economic Review 94 (1), 300–316 (March).

Kleinert, Jorn, Toubal, Farid, 2005. Gravity for FDI. Working Paper. Georg-August-Universitat Gottingen.

Lai, Huiwen, Zhu, Susan Chun, 2006. U.S. exports and multinational production. Review of Economics and Statistics 88 (3), 531–548 (August).

Lipsey, Robert E., 2002. Home and host country effects of FDI. NBER Working Paper, vol. 9293. October.
Markusen, James R., 1995. The boundaries of multinational enterprises and the theory of international trade. Journal of Economic Perspectives 9 (2), 169–189 (Spring).
Markusen, James R., 2002. Multinational Firms and the Theory of International Trade. MIT Press, Cambridge, MA.
Markusen, James R., Venables, Anthony J., 2000. The theory of endowment, intra-industry, and multi-national trade. Journal of International Economics 52, 209–234 (December).
Martin, Philippe, Rey, Helene, 2004. Financial supermarkets: size matters for asset trade. Journal of International Economics 64 (2), 335–361 (December).
McCulloch, Rachel, 1985. US direct foreign investment and trade. In: Erdilek, Asim (Ed.), Multinationals as Mutual Invaders: Intra-Industry Direct Foreign Investment. St. Martin's Press, New York, pp. 129–151.
McCulloch, Rachel, 1993. New perspectives on foreign direct investment. In: Froot, Kenneth (Ed.), Foreign Direct Investment. University of Chicago Press, Chicago, pp. 37–53.
Mundell, Robert, 1957. International trade and factor mobility. American Economic Review 47 (3), 321–335.
John, Mutti, Grubert, Harry, 2004. Empirical asymmetries in foreign direct investment and taxation. Journal of International Economics 62 (2), 337–358 (March).
Shim, Kieun, 2005. Border effects and FDI: can FDI explain the border puzzle? Working Paper. U. of Virginia.
Slaughter, Matthew J., 2000. Production transfer within multinational enterprises and American wages. Journal of International Economics 50 (2), 449–472 (April).