The *New Expats*: Economic Determinants of Bilateral Expatriate, FDI, and International Trade Flows

Jeffrey H. Bergstrand, Peter Egger, and Mario Larch

Vienna, OeNB, November 18, 2008

Before I start ...

Many thanks for having me!

And apologies for coming as a wolf in disguise …
with a migration brain-wash
[“my” migrants’ pockets and coffers are swelling of money]

I. Motivation

TIME magazine (October 22, 2007): "New Expatriates":

There is a swelling middle class of managers, engineers, and other professionals that MNEs move around the globe fluidly – much like capital and goods – to maximize production efficiency and economic profits.

II. Motivation

Friedman's (2005) *The World is Flat:*

"To begin with, he said, the notebook was co-designed in Austin, Texas, and in Taiwan by a team of Dell engineers and a team of Taiwanese notebook designers..... We put our engineers in their facilities and they come to Austin and we actually co-design these systems."

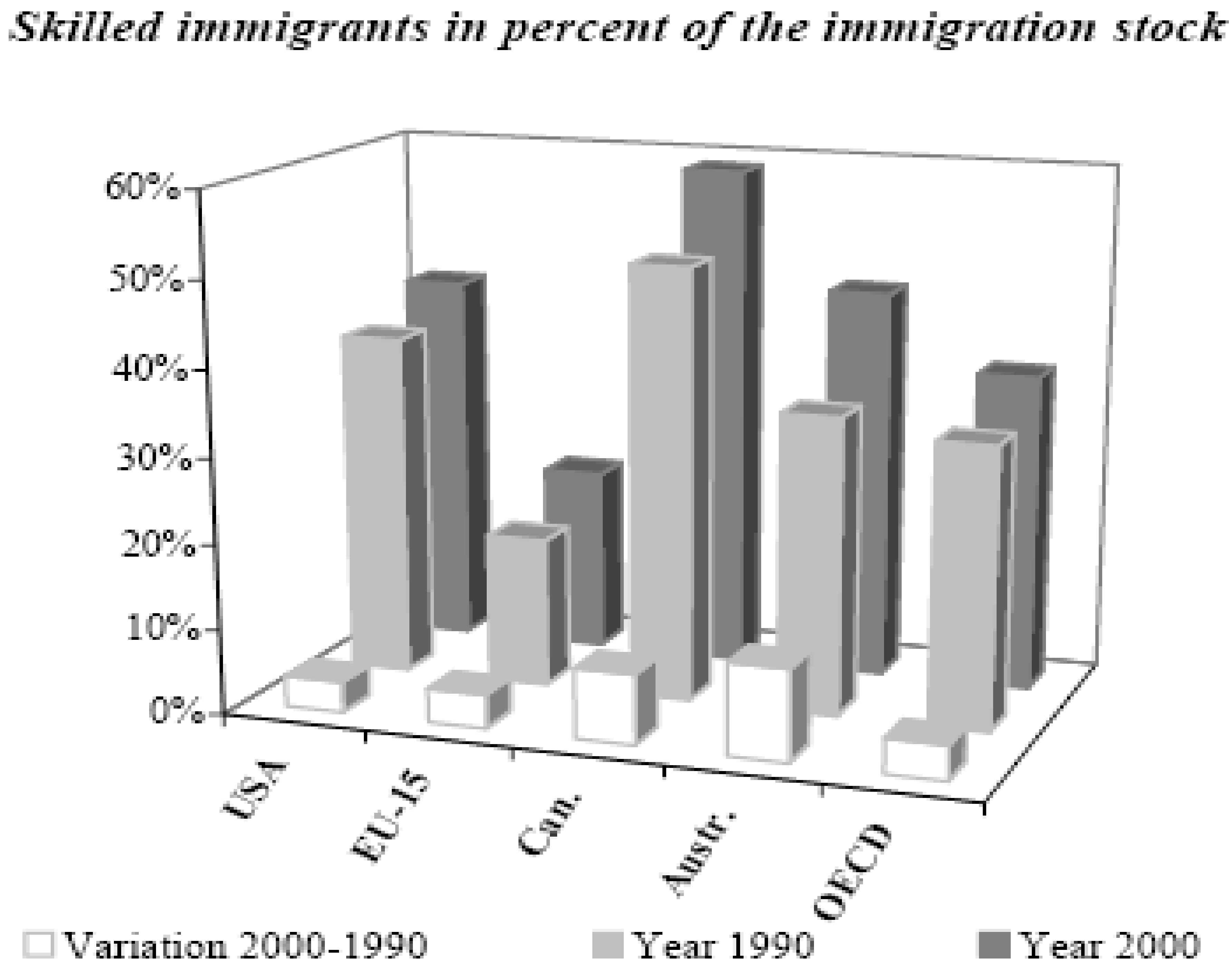
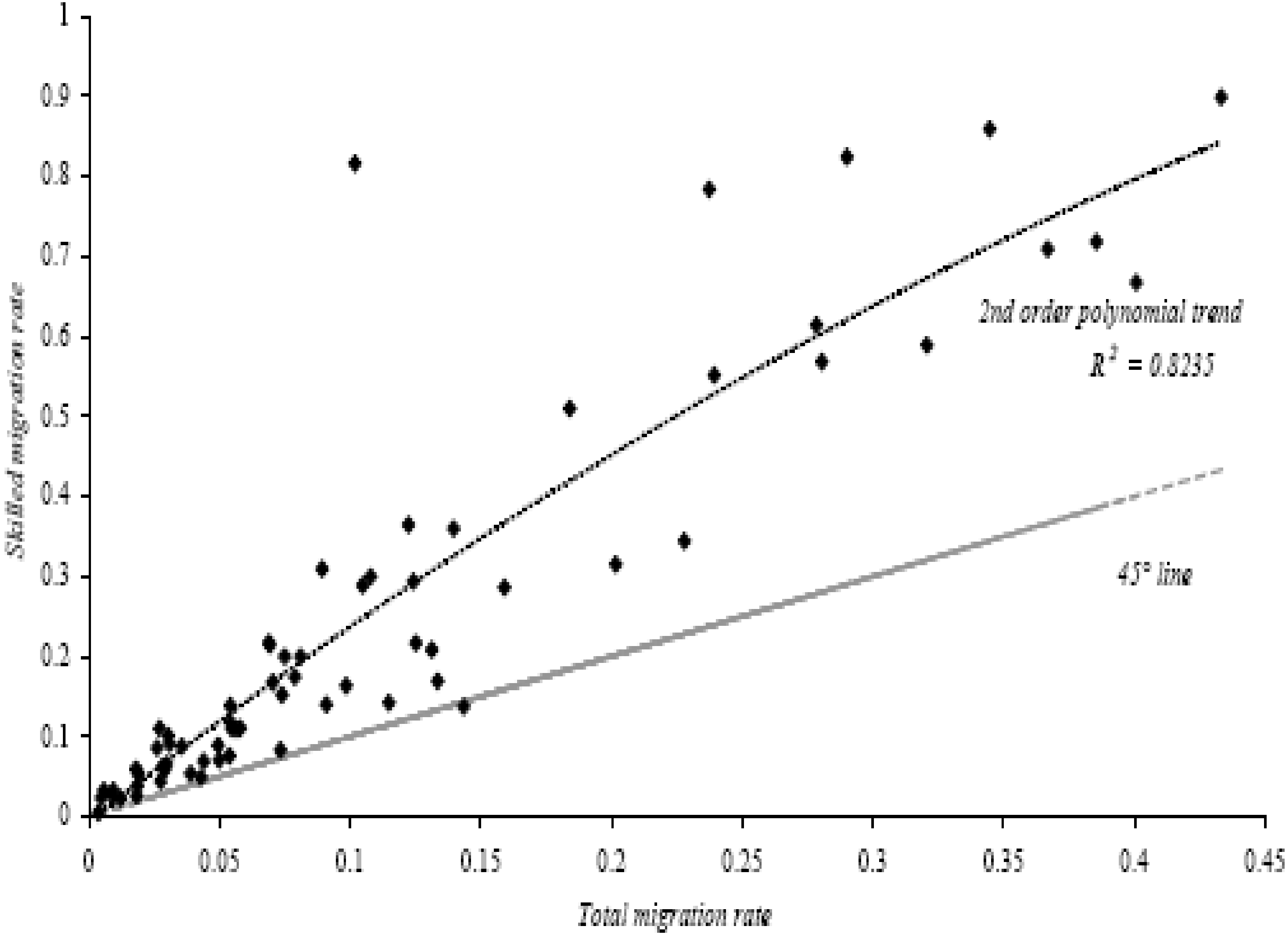

Skilled immigrants in percent of the immigration stock
60%
50%
40%
30%
20%
10%
0%
USA
EU-15
Can.
Austr.
OECD
Variation 2000-1990
Year 1990
Year 2000

Figure 1. Skilled migration rate and total emigration rate in 2000



III. Motivation

- New Expats:
  Labor-demand decisions of profit-maximizing MNEs.

- Traditional migration literature:
  - Unskilled labor migration.
  - Labor supply decision of utility-maximizing households.

IV. Motivation

The economic analysis of expats falls naturally into the domain of international trade and FDI flows for two reasons:

a) Supply of managers and professionals very elastic with respect to demand (Straubhaar and Wolter, 1997; Yeatman and Berden, 2007).

b) Data suggest two-way flows: Relative per capita incomes can only explain one-way flows.

Contributions

1. We develop a general-equilibrium-based theoretical framework to rationalize the existence of *two-way* expat flows as complements to (two-way) FDI and trade flows.
2. We motivate a theoretical rationale for estimating “gravity equations” of two-way expat flows.
3. We show that both horizontal *and* vertical MNE motives explain patterns of highly skilled migration.

Contribution 1

A 3x3x2 Model of Bilateral Trade, FDI/FAS, and Skilled Migration

a) 3 Factors:
   Unskilled labor, skilled labor, and physical capital (the latter two are imperfectly mobile internationally).
b) 3 Countries (i, j, ROW)
c) 2 Goods:
   Differentiated good X produced under IRS and monopolistic competition.
   Homogeneous good Y produced under CRS and perfect competition.

Utility and Production

Utility:

- Cobb-Douglas upper-tier utility function (X, Y).
- CES lower-tier function for varieties of X (from foreign national and vertical MNE exporters as well as local plants of MNE and NEs).

Production of X:

- CET between K and S.
- Cobb-Douglas between (K,S) and U.

The "Cross-Hauling" of Skilled Workers in Firms

i-based MNEs with plants in j use:

- δ_{ij}^{d} domestic S at headquarters in i.
- δ_{ij}^{f} foreign S at headquarters.
- ζ_{ij}^{d} domestic S abroad in j.
- ζ_{ij}^{f} foreign S abroad in j.

Managers/researchers are shuffled around within firms almost like goods.

Firms are assumed to maximize profits given the technologies and the demand relationships suggested above. Let c_{Xi} denote marginal production costs of differentiated good X in country i, then, the profit functions consequently are:

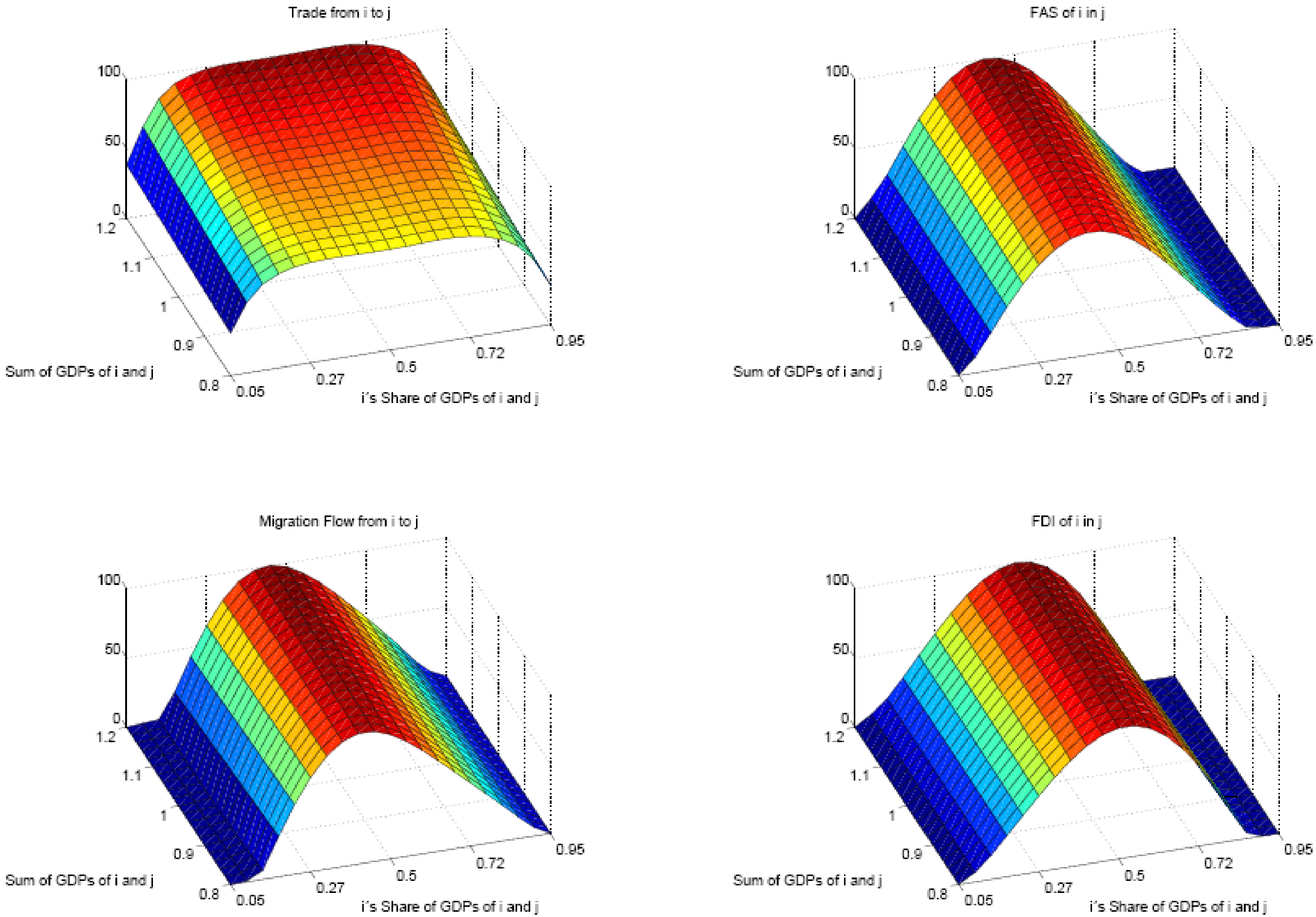
$$\pi^{n_i} = (p_{Xi} - c_{Xi}) \sum_{j=1}^{3} x_{ij} - w_{Si} - r_i. \tag{12}$$

$$\begin{aligned}\pi^{h_{3,i}} &= \sum_{j=1}^{3} (p_{Xj} - c_{Xj}) x_{jj} - \left[1 + \sum_{j \neq i} \delta_{ij}^{d} + \sum_{j \neq i} \zeta_{ij}^{d} \mu\right] w_{Si} - \sum_{j \neq i} \delta_{ij}^{f} \mu w_{Sj} \\ &+ \sum_{j \neq i} \zeta_{ij}^{f} w_{Sj} - \left[1 + \sum_{j \neq i} \gamma_{ij}^{d}\right] r_i - \left[\sum_{j \neq i} \gamma_{ij}^{f} r_j\right].\end{aligned} \tag{13}$$

$$\begin{aligned}\pi^{h_{2,ij}} &= (p_{Xi} - c_{Xi}) x_{ii} + (p_{Xj} - c_{Xj}) x_{jj} - \left[1 + \delta_{ij}^{d} + \zeta_{ij}^{d} \mu\right] w_{Si} - \left[\delta_{ij}^{f} \mu + \zeta_{ij}^{f}\right] w_{Sj} \\ &- \left[1 + \gamma_{ij}^{d}\right] r_i - \gamma_{ij}^{f} r_j.\end{aligned} \tag{14}$$

$$\begin{aligned}\pi^{v_{ij}} &= (p_{Xj} - c_{Xj}) \sum_{k=1}^{3} x_{jk} - \left[1 + \delta_{ij}^{d} + \zeta_{ij}^{d} \mu\right] w_{Si} - \left[\delta_{ij}^{f} \mu + \zeta_{ij}^{f}\right] w_{Sj} \\ &- \gamma_{ij}^{d} r_i - \gamma_{ij}^{f} r_j.\end{aligned} \tag{15}$$

Figures 3a-3f



Interpretation

In this model:

- FDI and expats are complementary.
- Trade and FDI/expats are substitutes or complements.

This suggests interesting conclusions about the influence of non-migration policy on expats.

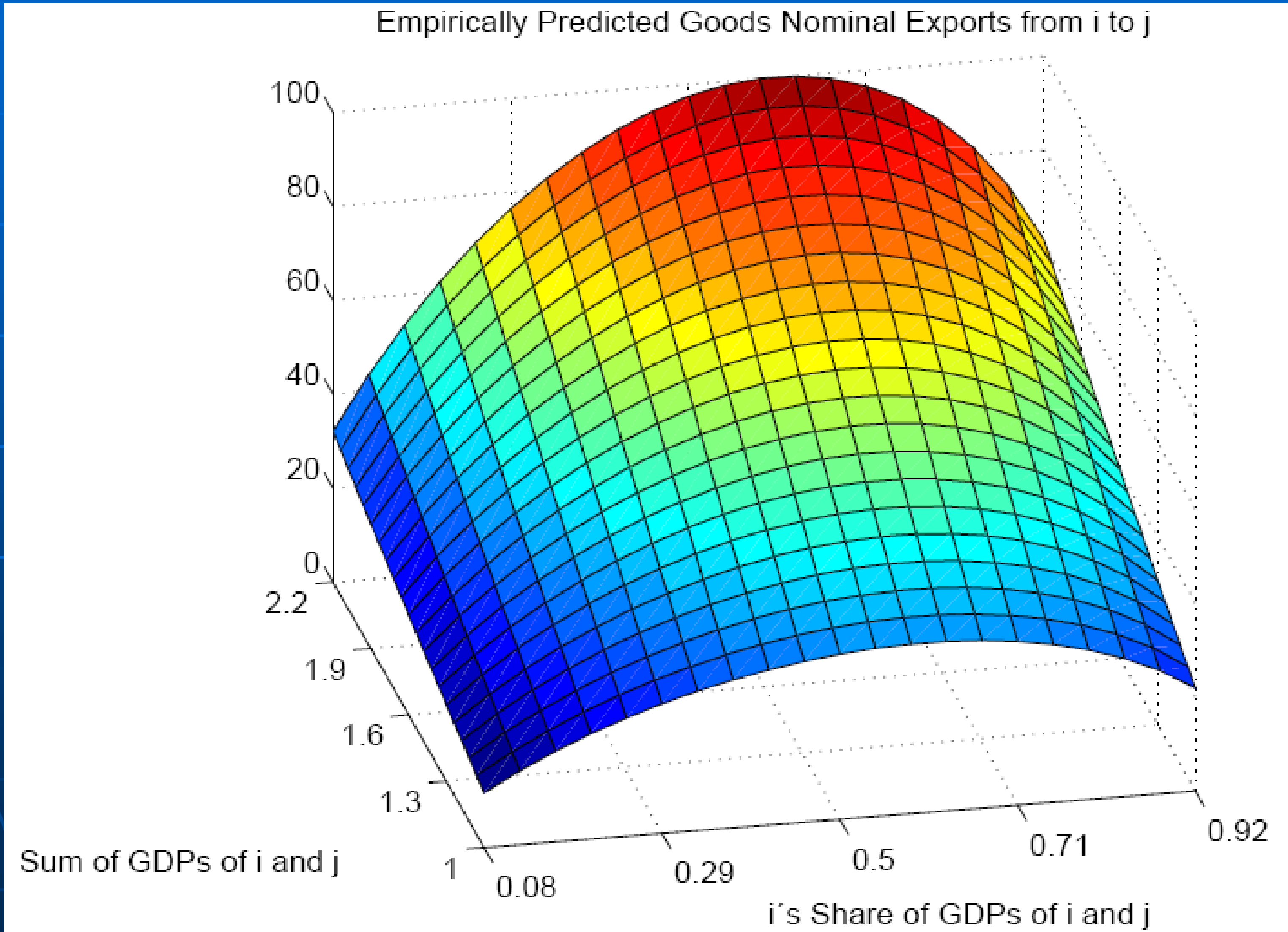
Contribution 2

- This model motivates gravity equations of trade, FDI, and expats.
- We estimate gross aggregate bilateral expatriate, trade, and FDI flows (using Poisson quasi-maximum likelihood; see Santos Silva and Tenreyro, 2006).

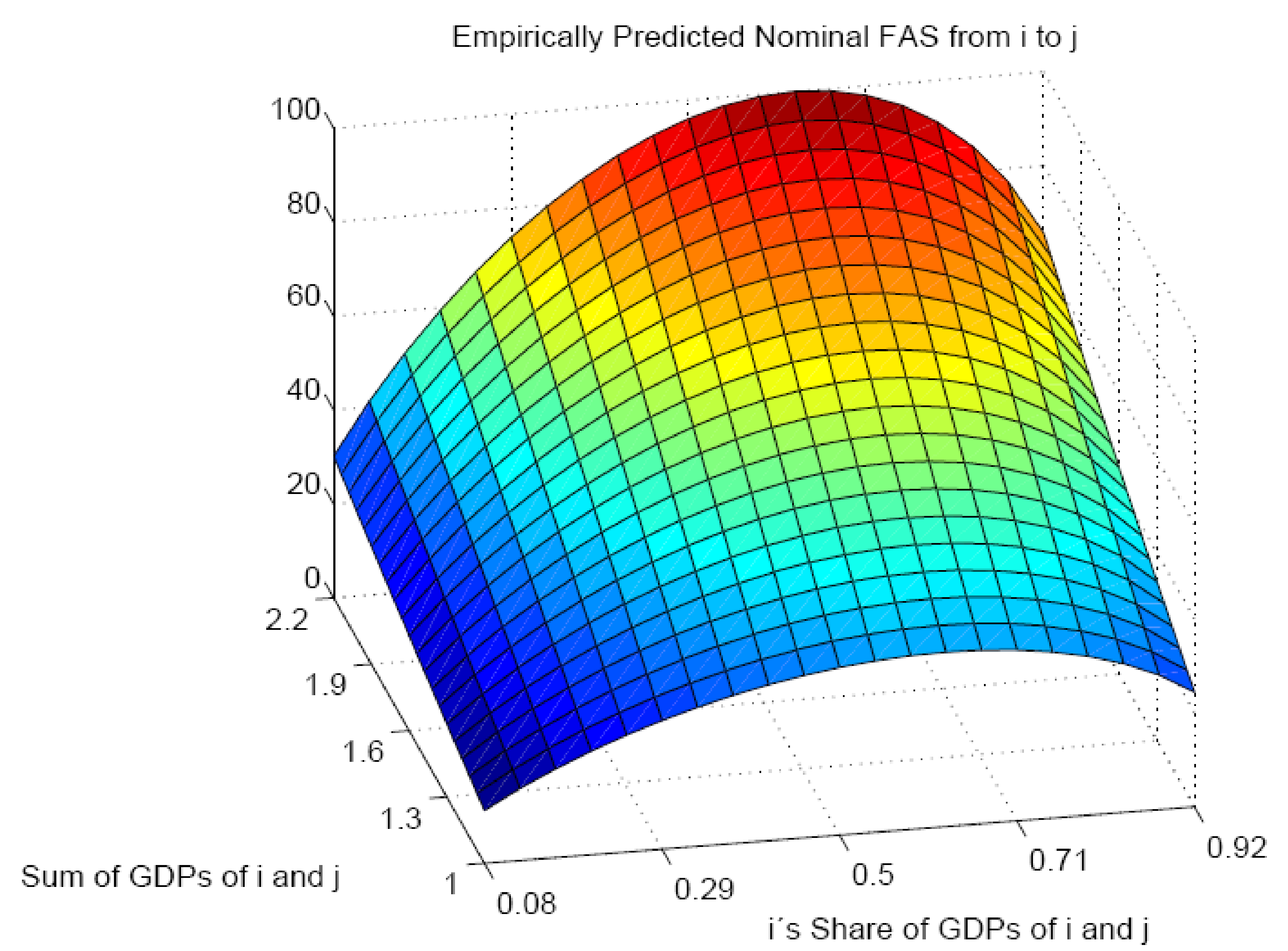
PPML Gravity Model Estimates

Variable	Exports	FAS	Expats
Exp. GDP	**0.777**	**0.787**	**0.591**
Imp. GDP	**0.516**	**0.496**	**0.880**
Distance	**-0.459**	**-0.405**	0.017
Adjacency	0.329	0.371	0.677
Language	0.171	0.222	**1.501**
Obs.	1152	1370	1925
Pseudo R^2	0.795	0.414	0.725

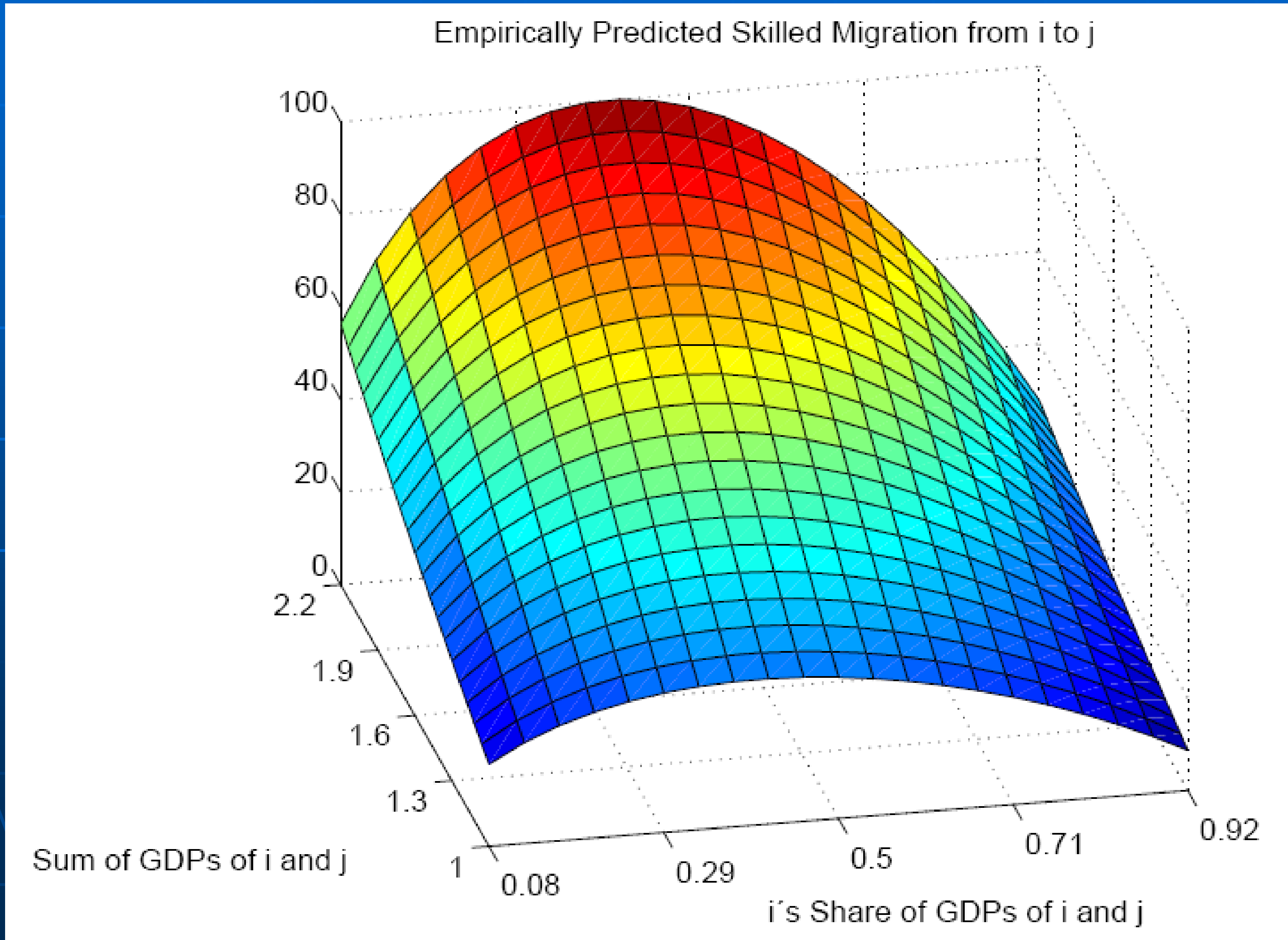
Graphical Illustration: Exports



Graphical Illustration: FAS



Graphical Illustration: Expats



Contribution 3

- We explain the primary factors driving skilled migration *beyond gravity:* notably, relative factor endowments and cost of migration.
- Two-way flows of migrants are maximized where horizontal MNEs dominate.

Conclusions

a) Theoretical rationale for estimating simultaneously gravity equations of bilateral trade, FDI/FAS, and skilled migration.

b) A considerable amount of the bilateral variation in such flows *beyond gravity* is explained by relative factor endowments (and the prevalence of horizontal MNEs).

c) The model provides a theoretical rationale for simultaneous *two-way* flows of highly skilled migrants both *intra-industry* ***and intra-firm.***

Conclusions

Relevant policy conclusions:

- Trade policy and, especially, investment policy is also migration policy (at least, for expats)!

- Impediments to FDI reduce the cross-border flows of skilled migrants
(and the pool of “ideas” and managing capacity in all economies).

Thank you for your attention !

The New Expats: Economic Determinants of Bilateral Expatriate, FDI, and International Trade Flows*

Jeffrey H. Bergstrand†, Peter Egger‡, and Mario Larch§

July 18, 2008

Abstract

Globalization has led to dramatic growth in two-way flows of highly skilled professionals – what *TIME* magazine recently termed the "New Expatriates." However, no study has attempted to explain empirically these migrations of highly skilled expats motivated by profit-maximizing multinational enterprises (MNEs), much less one based upon a formal theoretical foundation. We provide three potential contributions. First, we develop a general-equilibrium-based theoretical framework to rationalize the existence of *two-way* expat flows as complements to (two-way) foreign direct investment (FDI) and trade flows. Second, we motivate a theoretical rationale for estimating "gravity equations" of two-way expat flows, economic size, and economic similarity – and in a manner consistent with estimating gravity equations of bilateral trade, foreign affiliate sales, and FDI flows. Third, we show that – due to relative factor endowment differences – both horizontal *and* vertical MNE motives explain patterns of highly skilled migration flows; however, the bulk of such flows are explained by *two-way* expat flows among horizontal MNEs.

Key words: Expatriates, Foreign Direct Investment, Multinational Enterprises, International Trade, Gravity Equation

JEL classification: F1, F21, F22, F23

* *Acknowledgements*: To be added.

† *Affiliation:* Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, and CESifo Munich. *Address:* Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556 USA. *E-mail:* bergstrand.1@nd.edu.

‡ *Affiliation:* Ifo Institute for Economic Research, Ludwig-Maximilian University of Munich, CESifo, and Centre for Globalization and Economic Policy, University of Nottingham. *Address:* Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany. *E-mail:* egger@ifo.de.

§ *Affiliation:* Ifo Institute for Economic Research and CESifo. *Address:* Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany. *E-mail:* larch@ifo.de.

1 Introduction

> *"To begin with, he said, the notebook was codesigned in Austin, Texas, and in Taiwan by a team of Dell engineers and a team of Taiwanese notebook designers. . . . We put our engineers in their facilities and they come to Austin and we actually codesign these systems." (Thomas Friedman, The World is Flat, 2005, p. 415)*

In its October 22, 2007 issue, *TIME* magazine wrote about the "New Expatriates." Once the romantic domain of U.S. writers such as Ernest Hemingway and F. Scott Fitzgerald and the "exotic detour" of top multinational executives, the article noted that the "new expatriates" are a swelling *middle class* of managers, engineers, and other professionals that multinational enterprises move around the globe fluidly – much like capital and goods – to maximize production efficiency and economic profits. Such expatriates have actually been an important economic force since the dramatic rise in the number of multinational enterprises (MNEs) following World War II, generated by the labor-demand decisions of profit-maximizing MNEs. However, the traditional migration literature has paid scant attention to them – focusing instead upon unskilled migration flows generated by the labor supply decisions of utility-maximizing households, cf., Borjas (1994), Hatten and Williamson (2002), and Hanson (2006). However, to date, there has been *no systematic empirical analysis* of the economic determinants of world aggregate bilateral flows of such expatriates – much less one based upon a general equilibrium theory *or* fully consistent with the patterns of bilateral foreign direct investment (FDI) and international trade flows of profit-maximizing firms.

This paper provides the first integrated theoretical and empirical model explaining the patterns of bilateral aggregate expatriate, FDI, and international trade flows based on common economic determinants, with three potential contributions in mind. First, we develop a general-equilibrium-based theoretical framework to rationalize the existence of *two-way* expat flows as complements to (two-way) FDI and trade flows. Second, we motivate a theoretical rationale for estimating "gravity equations" of two-way expat flows, economic size, and economic similarity – and in a manner consistent with estimating gravity equations of bilateral trade, foreign affiliate sales, and FDI flows. Third, we show that – due to relative factor endowment differences – both horizontal *and* vertical MNE motives explain patterns of highly skilled migration flows; however, the bulk of such flows are explained by *two-way* expat flows among horizontal MNEs. This suggests that the "brain exchange" of highly skilled workers between countries of similar economic development tends to dominate the "brain drain" between countries of dissimilar economic development. We now summarize in turn each aspect.

First, the economic analysis of expatriates falls more naturally into the domain of international trade and FDI flows rather than the migration literature for two reasons. As just noted, the decision of a manager or professional working for an MNE to migrate abroad is largely based upon the labor-*demand* decision of the profit-maximizing MNE. As Straubhaar and Wolter (1997) note, the supply of managers and professionals within firms appears to be very elastic with respect to demand. These new expatriates see assignments abroad as a necessary condition (like an MBA degree) to advance their careers within MNEs, cf., Yeatman and Berden (2007). Consequently, the migration literature's emphasis on emigration determined by the labor-*supply* decision of households seems ill-suited as the major motivating force. Also, one of the most prominent economic variables explaining traditional unskilled and skilled migration flows is relative per capita incomes (or relative real wage rates), which can explain only *one-way* migration flows. However, the data suggest strong evidence of *two-way* flows of expatriates. Moreover, such flows tend to complement multinationals' FDI flows. The existence of two-way flows of highly skilled migrants that complement FDI requires an entirely different framework of analysis than the traditional migration literature provides.[1]

Since the role of MNEs is central to the analysis of expatriate flows, a theoretical framework to explain expatriate flows requires a role for multnational enterprises alongside

[1]For clarification, the study of *one-way* (inter-industry) trade has long been the domain of Ricardian and Heckscher-Ohlin models. "One-way" refers here to the outcome in the classical 2x2x2 trade model where one country exports, say, wine and the other country exports, say, cloth.

national enterprises (NEs). Consequently, our first goal is to develop a 3-country, 3-factor, 2-good general equilibrium model with internationally *mobile* skilled labor and physical capital (and internationally immobile unskilled labor) to motivate the economic determinants of two-way (alongside one-way) aggregate bilateral world trade, FDI, foreign affiliate sales (FAS), *and* expatriate flows. Markusen's (2002) 2x2x2 "knowledge-capital" model with two internationally *immobile* skilled and unskilled labor is a special case of our model. However, to capture the role of internationally mobile expatriates, we allow skilled labor to migrate (i.e., expatriates), where expatriations of managers are determined by the profit-maximizing decisions of labor-demanding MNEs. The 3-factor, 3-country, 2-good knowledge-and-physical-capital model in Bergstrand and Egger (2007) is also a special case of our model. One of the important features of 2-factor knowledge-capital-type models is that they can explain simultaneously international trade flows and FAS; however, they cannot explain FDI flows, cf., Markusen (2002, p. 8). Bergstrand and Egger (2007) introduced a third factor (internationally mobile physical capital) to motivate FDI as well as FAS, and introduced a third country (ROW) to motivate bilateral trade flows in a multi-country world.[2] Our model here rationalizes formally the existence of two-way expatriate flows, and effectively generates a theory of *intra-industry* – as well as inter-industry – trade, FDI, FAS, and expatriates simultaneously.[3]

Second, just as the knowledge-capital literature has tended to ignore migration flows, the literature on international migration has tended to ignore trade and FDI.[4] Very early analyses of bilateral aggregate migration flows actually used gravity equations to explain

[2]See Bergstrand and Egger (2007) on the distinctions between the terms FDI and FAS, noting now that these are two different "concepts" of MNE behavior.

[3]The literature just discussed focuses on determinants of aggregate bilateral trade and FAS flows, and does not address the recent literature assuming exogenous heterogeneous productivities to explain (in data at the firm and plant levels) *which firms* select into national firms versus MNEs, as discussed in Helpman (2006). For tractability, this issue is beyond the scope of this paper. Also, the Knowledge-Capital literature typically assumes that the "services" of skilled labor move costlessly between headquarters and foreign affiliates. The existence of expatriates in reality implies that the flow of services of such skilled workers are not perfectly costless.

[4]There have been several empirical studies that have examined the "effect" of bilateral FDI or migration flows on trade, as well as the effect of trade on FDI and migration flows. However, these approaches suggest that all three flows are *not* determined simultaneously in the long-run by common economic factors.

such flows but lacked formal theoretical foundations, cf., Ravenstein (1885, 1889) and other references cited later. Even though gravity equations explain bilateral migration flows empirically very well (as such equations explain bilateral trade and FDI flows well), the modern literature on migration – motivated more by the labor-supply decision of income-maximizing workers – has focused instead on per capita income (or wage) differentials and on policy impediments to migration to explain emigration from the labor supply side, cf., Borjas (1994, 1999). In the absence of idiosyncratic factors (e.g., family considerations) that might influence the migration decision of individuals, these models based upon relative wage rates can only explain *one-way* bilateral aggregate migration flows, not *two-way* flows – which are observed prominently for skilled migrants.[5]

Our second goal then is to use our general equilibrium model to provide a theoretical rationale for estimating gross bilateral two-way flows of highly skilled migrants using the *gravity equation*. Since the modern migration literature's focus on income-maximizing emigration-supply decisions of workers precludes motivating a gravity equation for bilateral aggregate migration flows because of the focus on explaining one-way flows, we use our theory of two-way emigration of highly skilled expatriates to motivate estimating aggregate expatriate flows using the gravity equation's multiplicative form. While theoretical foundations for international trade gravity equations are now well accepted, until Bergstrand and Egger (2007) there was no logically-consistent theoretical foundation for FDI flows – much less one that could explain simultaneously FDI *and* trade flows. In this paper, we provide a theoretical foundation for estimating gravity equations of gross aggregate bilateral expatriate, trade, and FDI flows. Moreover, the theoretical model also suggests *quantitatively different* gravity "relationships" of the flows with economic size and similarity.

The opportunity to explore empirically bilateral highly skilled migration flows, including "expats," has been made possible only recently because of the construction of a new OECD data set. OECD (2005) provides data on bilateral aggregate migration flows by

[5]Similarly, the economic geography literature uses an admittedly "*ad hoc*" specification to model only one-way migration flows using wage differentials; we provide more detail shortly.

country for each OECD country with approximately 100 countries, separating migrants by primary-education (unskilled), secondary-education (more skilled), and tertiary-education or higher (highly skilled) levels. Approximately 50 percent of migration flows of OECD countries are *skilled* flows (more skilled and highly skilled), cf., Figure 1a from Docquier and Marfouk (2004). Moreover, for the vast bulk of countries in the world, skilled migration is growing at *twice* the rate of unskilled migration, as implied by Figure 1b from Docquier and Marfouk (2004) using a cross-section of skilled emigration rates (on y-axis) vs. total emigration rates (on x-axis). In fact, two-thirds of skilled emigrants of North America and Europe are tertiary-level educated, and the largest group of highly skilled migrants from the European Union (EU) to the United States (US) – 81 percent – are *executives and managers*. We provide empirical evidence using a cross-section of two-thousand country pairs in year 2000 that the bilateral flows of highly skilled migrants can be explained very well by a gravity equation, but *not the same* gravity relationships as for trade and FDI flows. But we find that our *empirical* gravity equations are consistent with their respective *theoretical* gravity equations.[6]

Our third goal is to use our model to examine economic factors influencing expat flows *beyond gravity* – relative factor endowments, alongside costs of migration. Because of numerous nonlinear interaction terms between factor endowments and other right-hand-side (RHS) variables in our empirical specifications, we turn to a standard tool of trade economists – the Edgeworth box – to analyze in a tractable fashion the theoretical and empirical relationships between relative factor endowments, relative wage rates, and migration flows of highly skilled workers. The theoretical model predicts relationships between relative factor endowments and expat flows, illustrated with an Edgeworth box. Using the empirically-fitted values from Poisson Quasi-Maximum Likelihood estimation, we can also plot the empirically-predicted (highly skilled) migration flows against observed relative factor endowments with an Edgeworth box. The theoretical model predicts the

[6]Our particular OECD data set focuses explicitly on tertiary-level-educated (or higher) migrants. In reality, some of the highly skilled migrants in our sample are likely motivated by traditional income-maximizing models of workers; however, adding a conceptual analysis of such migrants as well is beyond the scope of this paper and left for future research.

"empirically-predicted" flows remarkably well and the Edgeworth-box approach allows for tractable economic interpretation of the empirical results. Moreover, by allowing for relative factor endowment differences alongside absolute factor endowment differences, we can actually decompose *quantitatively* the relative influences of *horizontal* versus *vertical* MNE motivations in the bilateral migration flows. Although horizontal and vertical MNE motivations *both* matter for explaining highly skilled migration flows, we show that the bulk of the variation in expat flows cross-sectionally can be attributed to *two-way* flows between horizontal MNEs, or the so-called "brain exchange."

Finally, the migration literature has shown a positive correlation between skilled emigration rates and relative skilled wage rates across country pairs, that is, an observed "positive sorting" of higher skilled workers into destination countries with higher rewards to skilled-*relative-to*-unskilled labor (cf., Hatton and Williamson, 2005; Grogger and Hanson, 2008).[7] Because factor prices and factor flows are both endogenous in our theoretical model, we do not try to show that relative skilled wage differences across countries (or relative returns to skills within destination countries) *cause* emigrations of highly skilled. However, we can show that our theoretical model predicts the observed positive correlation between highly skilled migration flows and relative returns to skills within destination countries found in the migration literature, that is, "positive sorting" and the "brain drain."[8]

The remainder of this paper is as follows. Section 2 presents a theoretical knowledge-and-physical-capital model with highly skilled migrants (expats). Section 3 provides an explanation of the calibration of the numerical version of our model. In section 4, we provide a theoretical rationale for two-way expatriate flows and for estimating simultaneously

[7]The other important feature of the migration literature noted in Grogger and Hanson (2008) is "positive selection," meaning that most immigrations bring more skilled workers into a country. We do not explain "positive selection" because of the model's assumption of immobile unskilled labor; our model can only at this time allow skilled migrants.

[8]The Economic-Geography literature, based upon the "Core-Periphery" (C-P) model of Fujita, Krugman, and Venables (1999), addresses skilled migration, but cannot explain two-way skilled migration. The reason is that, in the C-P model, "migration is governed by an *ad hoc* migration equation," with skilled migration driven by the real wage rate difference (Baldwin et al., 2003, p. 2-6). In fact, unlike our model, the C-P model does not generate migration in long-run equilibrium.

gravity equations for bilateral trade, FDI/FAS, and (highly) skilled migration flows, and then we provide empirical support for these predictions. In section 5, we provide theoretical predictions for the relationships between bilateral skilled migration flows and relative factor endowments, and then we provide empirical support for these predictions. Section 6 provides a theoretical and empirical sensitivity analysis, illustrating the *interconnections* of trade, FDI/FAS, and skilled migration. Section 7 concludes.

2 Theoretical Framework

In this section, we develop a theoretical general equilibrium model to explain the existence and economic determinants of two-way (alongside one-way) bilateral highly skilled migration flows – in a manner that can explain simultaneously economic determinants of one-way and two-way bilateral trade, FDI, and foreign affiliate sales (FAS) flows. Since profit-maximizing MNEs are the driving (demand-side) factor in determining expatriate flows, their role is essential. We describe a 2-good, 3-factor, 3-country theoretical general equilibrium model of national and multinational enterprises with internationally mobile physical capital and skilled labor ("expats"), but immobile unskilled labor. The model we build is in the spirit of the "knowledge-capital" model of Markusen (2002) and the "knowledge-and-physical-capital" model of Bergstrand and Egger (2007, 2008); in fact, the latter two models are special cases of our model.

As background, the knowledge-and-physical-capital (KAPC) model in Bergstrand and Egger (2007, 2008) is a 3-factor, 3-country, 2-good extension of Markusen's 2x2x2 knowledge-capital (KC) model with national enterprises (NEs), horizontal multinational enterprises (HMNEs), and vertical multinational enterprises (VMNEs). The demand side in the KAPC model is analogous to that in the KC model. However, the KAPC model extends the KC model in two significant ways. The first distinction is to use three primary factors of production: unskilled labor, skilled labor, and physical capital. In the KAPC model, 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 U.S. Bureau of Economic Analysis definition of foreign "direct investment positions" using domestic and foreign-affiliate shares of real fixed investment. The existence of imperfectly mobile physical capital (FDI) allows FDI, FAS, and trade to coexist, even for two countries (i, j) that are identical in all respects. The second distinction of the KAPC model from the KC model is to introduce a third country (ROW), which motivates a theoretical rationale for estimating gravity equations of bilateral trade and FDI simultaneously.

The model in this paper is more general than the KC and KAPC models. In addition to having a third factor (physical capital) and allowing FDI, it allows profit-maximizing MNEs to also choose the optimal allocation of imperfectly-mobile skilled labor between home and foreign locations of headquarters and of plants. The globalization of labor markets takes various forms. For unskilled workers, emigration is likely influenced predominantly by *supply* considerations: workers in country i look at relative real wage rates between i and j (relative to the costs of migration) in determining their income-maximizing behavior, cf., Borjas (1994). However, as Straubhaar and Wolter (1997) note, for highly skilled workers at the other extreme, potential emigrants need to assess "the international transferability of their firm-specific and location-specific knowledge" (p. 174). They also note that intra-firm mobility of highly skilled workers in MNEs "proves a very efficient strategy for transferring firm-specific know-how from the headquarters to the subsidiaries and vice-versa" (p. 175). This literature notes two prominent features determining the supply of highly skilled workers for intra-firm mobility. First, evidence suggests that international *intra-firm* mobility has become relatively less costly for labor-market adjustment than inter-firm mobility. Second, to encourage a ready supply of highly skilled migrants, labor-demanding MNEs have generally compensated workers with higher salaries or promises of higher future earnings and promotions (i.e., a premium). Straubhaar and Wolter (1997) conclude, "As a consequence, it can be assumed that the firms' demand for employees willing to work abroad is *generally met* by a corresponding supply" (p. 178; italics added). In our model, domestic-headquartered MNEs can potentially use some domestic skilled employees to help setup plants abroad, or use some skilled foreign nationals to create a headquarters at home; however, there are costs to migration.[9]

2.1 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 (1977) 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), two-plant (h_2) or three-plant (h_3) horizontal multinational firms, 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 equation (1) below:

$$\begin{aligned} V_i \;=\; & \left[\sum_{j=1}^{3} n_j \left(\frac{x_{ji}^{n}}{1+\tau_{Xji}} \right)^{\frac{\varepsilon}{\varepsilon-1}} + \left(\sum_{j=1}^{3} h_{3,j} \left(x_{ii}^{h_3} \right)^{\frac{\varepsilon}{\varepsilon-1}} + \sum_{j \neq i} \left(h_{2,ij} + h_{2,ji} \right) \left(x_{ii}^{h_2} \right)^{\frac{\varepsilon}{\varepsilon-1}} \right) \right. \\ & \left. + \sum_{k \neq j} \sum_{j=1}^{3} v_{kj} \left(\frac{x_{ji}^{v}}{1+\tau_{Xji}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \right]^{\frac{\varepsilon-1}{\varepsilon}\eta} \left[\sum_{j=1}^{3} Y_{ji} \right]^{1-\eta}. \qquad (1) \end{aligned}$$

[9]For ease of terminology, we will generally refer to unskilled and skilled workers only, keeping in mind that "skilled" corresponds in our theoretical and empirical contexts technically to "highly skilled" (tertiary-educated) workers. We note that the model can potentially be extended in other directions as well. For instance, we assume homogenous productivities for national exporting firms and MNEs; heterogenous productivities addressed in Helpman (2006) are beyond the scope of the present paper, but are a potential issue to incorporate in future work.

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 firms in industry X sold to country i, n_j is the (endogenous) number of these national firms in j, and τ_{Xji} are resource-consuming trade costs from j to i of the iceberg type (expressed as a fraction) and detailed later.

The second set of components reflects *horizontal* multinational firms that may have plants in either two or three countries to be "proximate" to markets to avoid trade costs. HMNEs cannot export goods. Every HMNE has a plant in its headquarters country. Let $h_{3,j}$ denote the (endogenous) number of multinationals that produce in all three countries and are headquartered in j ($j = 1, 2, 3$), $h_{2,ij}$ denote the number of two-country multinationals headquartered in i with a plant also in j, and $h_{2,ji}$ denote the number of two-country multinationals headquartered in j with a plant also in i. $x_{ii}^{h_3}$ is output produced in country i (and consumed in i) of a three-country HMNE and $x_{ii}^{h_2}$ is the output produced in country i (and consumed in i) of a two-country multinational firm. Note that h_2 plants arise when market size in one of the three countries is inadequate to warrant a local plant, and is more efficiently served (given trade, investment, and migration costs) by its own national firms and imports from foreign firms.

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 (2007) and Blonigen, Davies, Waddell, and Naughton (2007). 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 $1 + t_{Xji}$ $(1 + t_{Yji})$ denote the gross trade cost for shipping differentiated (homogeneous) good X (Y) from j to i.[10] Let $1 + t_{Xji} = 1$ for $i = j$, and analogously for $1 + t_{Yji}$. It will be useful to define:

$$1 + t_{Xji} = (1 + b_{Xji})(1 + \tau_{Xji}), \qquad 1 + t_{Yji} = (1 + b_{Yji})(1 + \tau_{Yji}),$$

where 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 percent) 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:

$$\begin{aligned} E_i &= \sum_{j=1}^{3} n_j p_{Xj}^n (1 + b_{Xji}) x_{ji}^n + \sum_{j=1}^{3} h_{3,j} p_{Xi}^{h_3} x_{ii}^{h_3} + \sum_{j \neq i} \Big(h_{2,ij} + h_{2,ji} \Big) p_{Xi}^{h_2} x_{ii}^{h_2} \\ &+ \sum_{k \neq j} \sum_{j=1}^{3} v_{kj} p_{Xj}^v (1 + b_{Xji}) x_{ji}^v + \sum_{j=1}^{3} p_{Yi} Y_{ji}, \end{aligned}$$

where E_i is GDP (and total factor income; E_i is determined in Section 2.4) in country i; p_{Xi}^n, p_{Xi}^h, p_{Xi}^v, denote the "mill" prices charged by producers in i for goods X (NEs, HMNEs, and VMNEs, respectively); and p_{Yi} is the price consumers in i have to pay for good Y.

Maximizing (1) subject to the budget constraint yields the domestic demand functions:

$$x_{ii}^{\ell} = \left(p_{Xi}^{\ell} \right)^{\varepsilon - 1} P_{Xi}^{-\varepsilon} \eta E_i; \quad \ell = \{n, h_2, h_3, v\},$$

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

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

$$
\begin{aligned}
P_{Xi} \quad = \quad & \left[\sum_{j=1}^{3} n_j \left([1 + t_{Xji}] p^{n}_{Xj} \right)^{\varepsilon} + \sum_{j=1}^{3} h_{3,j} \left(p^{h_3}_{Xi} \right)^{\varepsilon} + \sum_{j \neq i} \Big(h_{2,ij} + h_{2,ji} \Big) \left(p^{h_2}_{Xi} \right)^{\varepsilon} \right. \\
& \left. + \quad \sum_{k \neq j} \sum_{j=1}^{3} v_{kj} \left([1 + t_{Xji}] p^{n}_{Xj} \right)^{\varepsilon} \right]^{\frac{1}{\varepsilon}} \qquad (2)
\end{aligned}
$$

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:

$$
\frac{x_{ji}}{x_{ii}} = \left(\frac{p_{Xj}}{p_{Xi}} \right)^{\varepsilon - 1} (1 + t_{Xji})^{\varepsilon} (1 + b_{Xji})^{\varepsilon - 1}. \qquad (3)
$$

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:

$$
\sum_{j=1}^{3} Y_{ji} = \frac{1 - \eta}{p_{Yi}} E_i. \qquad (4)
$$

2.2 Differentiated Goods Producers

As in Markusen (2002), we assume that all manufacturing firms face variable marginal costs of production and fixed costs of setups of headquarters and plants. We assume that manufactures can be produced in all three countries using skilled labor, unskilled labor, and physical capital. Each country i is assumed to be endowed with an exogenous amount of internationally immobile unskilled labor, U_i, whose price is w_{Ui}.

Each country i is endowed with an exogenous amount of skilled labor, S_i, whose price is w_{Si}. Yet, in contrast to unskilled workers, skilled workers may be used abroad within an MNE to setup a foreign affiliate plant (alongside foreign skilled labor) and hence move across borders, depending upon skilled labor prices. These “expatriates” are rewarded their home factor price, plus an exogenous premium (or migration cost), but they consume in the country of residence (host). Also, a domestic MNE may require some skilled foreign nationals to create a viable multinational firm, depending upon skilled labor prices. The ability of MNEs to draw on domestic and foreign skilled labor markets to establish headquarters and plants allows for an endogenous determination of two-way bilateral flows of skilled workers in our 3-country model, consistent with empirical observations. We will see later that the composition of skilled workers (domestic vs. foreign) for the creation of headquarters and foreign plants by an MNE will depend on the domestic vs. foreign price of skilled workers.

Each country i is also endowed with an exogenous amount of physical capital, K_i, whose price is r_i. Similar to skilled labor, physical capital can be transferred endogenously across countries to setup plants by MNEs to maximize their profits, thus making for an endogenous determination of bilateral FDI flows. Analogous to their usage of skilled workers, MNEs use a composite of domestic and foreign physical capital for the setup of plants, and an MNE's determination of foreign versus domestic capital will depend on the relative capital rental rates.

Differentiated goods producers operate in monopolistically competitive markets, similar to Markusen (2002, Ch. 6). 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.[11] 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:

$$F_{Xi} = (K_{Xi}^{\chi} + S_{Xi}^{\chi})^{\frac{\alpha}{\chi}} \, U_{Xi}^{1-\alpha}. \tag{5}$$

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

[11] As mentioned earlier, we do not assume heterogeneous productivities, but the model could be extended in the future to allow this.

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 is in Goldin and Katz (1998) and (in the MNE literature) in Slaughter (2000). We nest a CES production function within the Cobb-Douglas function to allow for potential complementarity of physical and human capital in production; χ determines the degree of complementarity or substitutability.

The presence of fixed costs allows a rich opportunity to provide a theoretical rationale for skilled migration and FDI flows. First, we assume that every firm (NE or MNE) faces an exogenous requirement of one unit of home skilled labor to establish a headquarters in the home country (say, i), the cost of which to the firm is w_{Si}. Also, every plant at home requires exogenously one unit of home physical capital for setup, the cost of which is r_i. Hence, a single-plant national firm in i that exports abroad faces fixed costs of $w_{Si}+r_i$.

Second, NEs and MNEs differ in fixed costs. Each MNE incurs one headquarter setup, the cost of which is assumed larger than that of an NE, as in Markusen (2002). We assume that an MNE's headquarters setup requires $1+\delta$ units of skilled labor ($\delta > 0$).

Third, and unique to this paper, we assume that profit-maximizing MNEs in country i can (potentially) use foreign skilled labor as well as domestic skilled labor to setup their headquarters. For instance, many MNEs use foreign nationals in headquarters to provide essential "knowledge" of the foreign market, depending upon skilled labor prices abroad. This creates "expatriates" from country j and ROW (our third country). Hence, δ can be non-linearly decomposed into δ_{ij}^d for domestic skilled labor and δ_{ij}^f for foreign skilled labor. To introduce relative wage costs of domestic and foreign skilled labor into the MNE's profit-maximizing decision, we assume a simple constant-elasticity-of-transformation (CET) technology where δ_{ij}^d and δ_{ij}^f are given by:

$$\delta_{ij}^d = w_{Si}^{\frac{1}{\rho_\delta - 1}} \psi_i^{\frac{-\rho_\delta}{\rho_\delta - 1}} \left(w_{Si}^{\frac{\rho_\delta}{\rho_\delta - 1}} \psi_i^{\frac{-\rho_\delta}{\rho_\delta - 1}} + (w_{Sj}\mu)^{\frac{\rho_\delta}{\rho_\delta - 1}} \psi_j^{\frac{-\rho_\delta}{\rho_\delta - 1}} \right)^{\frac{-1}{\rho_\delta}} \delta, \tag{6}$$

$$\delta_{ij}^{f} = (w_{Sj}\mu)^{\frac{1}{\rho_\delta - 1}} \psi_j^{\frac{-\rho_\delta}{\rho_\delta - 1}} \left(w_{Si}^{\frac{\rho_\delta}{\rho_\delta - 1}} \psi_i^{\frac{-\rho_\delta}{\rho_\delta - 1}} + (w_{Sj}\mu)^{\frac{\rho_\delta}{\rho_\delta - 1}} \psi_j^{\frac{-\rho_\delta}{\rho_\delta - 1}} \right)^{\frac{-1}{\rho_\delta}} \delta, \tag{7}$$

where μ is the (gross) premium on high-skilled expatriates' domestic wages, which introduces a migration "cost." We assume that the premium is exogenous and homogeneous across countries and firms. $\psi_i^{\frac{-\rho_\delta}{\rho_\delta - 1}}$ $\left(\psi_j^{\frac{-\rho_\delta}{\rho_\delta - 1}}\right)$ is a weight attributed to domestic (foreign) wages w_{Si} (w_{Sj}), and ρ_δ determines the elasticity of substitution between domestic and foreign skilled labor for setting up the MNE headquarters in i. For a 2-country HMNE or a VMNE with headquarters in i and a plant in j, $\delta = \left[\left(\psi_i \delta_{ij}^{d}\right)^{\rho_\delta} + \left(\psi_j \delta_{ij}^{f}\right)^{\rho_\delta} \right]^{1/\rho_\delta}$. For 3-plant HMNEs of i, $\delta = \frac{1}{2} \left\{ \left[\left(\psi_i \delta_{ij}^{d}\right)^{\rho_\delta} + \left(\psi_j \delta_{ij}^{f}\right)^{\rho_\delta} \right]^{1/\rho_\delta} + \left[\left(\psi_i \delta_{iROW}^{d}\right)^{\rho_\delta} + \left(\psi_{ROW} \delta_{iROW}^{f}\right)^{\rho_\delta} \right]^{1/\rho_\delta} \right\}$, where δ_{iROW}^{d} and δ_{iROW}^{f} are defined analogously.

Fourth, and also unique to this paper, we assume that profit-maximizing MNEs can (potentially) use domestic skilled labor as well as foreign skilled labor to setup a foreign affiliate. This creates "expatriates" for country i. In reality, foreign plants of MNEs need some skilled labor at the foreign location to setup an affiliate abroad. Our framework assumes that the setup of any plant abroad requires $1 + \zeta$ units of skilled labor, compared with the 1 unit of home skilled labor for an NE to setup a plant at home (see above). To similarly introduce relative wage costs of domestic and foreign skilled labor into the MNE's profit-maximizing decision, we also assume a simple CET technology where ζ_{ij}^{d} and ζ_{ij}^{f} are given by:

$$\zeta_{ij}^{d} = (w_{Si}\mu)^{\frac{1}{\rho_\zeta - 1}} \psi_i^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} \left((w_{Si}\mu)^{\frac{\rho_\zeta}{\rho_\zeta - 1}} \psi_i^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} + w_{Sj}^{\frac{\rho_\zeta}{\rho_\zeta - 1}} \psi_j^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} \right)^{\frac{-1}{\rho_\zeta}} \zeta, \tag{8}$$

$$\zeta_{ij}^{f} = w_{Sj}^{\frac{1}{\rho_\zeta - 1}} \psi_j^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} \left((w_{Si}\mu)^{\frac{\rho_\zeta}{\rho_\zeta - 1}} \psi_i^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} + w_{Sj}^{\frac{\rho_\zeta}{\rho_\zeta - 1}} \psi_j^{\frac{-\rho_\zeta}{\rho_\zeta - 1}} \right)^{\frac{-1}{\rho_\zeta}} \zeta, \tag{9}$$

where $\psi_i^{\frac{-\rho_\zeta}{\rho_\zeta - 1}}$ $\left(\psi_j^{\frac{-\rho_\zeta}{\rho_\zeta - 1}}\right)$ is a weight attributed to domestic (foreign) wages w_{Si} (w_{Sj}), and ρ_ζ determines the elasticity of substitution between domestic and foreign skilled labor for setting up a plant in j. For a 2-country HMNE or a VMNE with headquarters in i and a plant in j, $\zeta = \left[\left(\psi_i \zeta_{ij}^d\right)^{\rho_\zeta} + \left(\psi_j \zeta_{ij}^f\right)^{\rho_\zeta}\right]^{1/\rho_\zeta}$. For 3-plant HMNEs of i, $\zeta = \frac{1}{2}\left\{\left[\left(\psi_i \zeta_{ij}^d\right)^{\rho_\zeta} + \left(\psi_j \zeta_{ij}^f\right)^{\rho_\zeta}\right]^{1/\rho_\zeta} + \left[\left(\psi_i \zeta_{iROW}^d\right)^{\rho_\zeta} + \left(\psi_{ROW} \zeta_{iROW}^f\right)^{\rho_\zeta}\right]^{1/\rho_\zeta}\right\}$, where ζ_{iROW}^d and ζ_{iROW}^f are defined analogously. Together, δ and ζ determine the additional requirement of skilled labor for setting up an MNE as compared to an NE.

Fifth, we assume that profit-maximizing MNEs can (potentially) use domestic and foreign physical capital to setup plants abroad, similar to Bergstrand and Egger (2007, 2008). This creates "FDI" from country i to country j. Our framework assumes that the setup of any plant abroad to initiate production and sales requires $1+\gamma$ units of physical capital, compared with 1 unit of physical capital for an NE or MNE to set up a plant at home. Thus, γ effectively embeds "investment costs" into the model. Like above, to introduce relative costs of domestic and foreign physical capital into the MNE's profit-maximizing decision, we again assume a simple CET function where γ_{ij}^d and γ_{ij}^f are:

$$\gamma_{ij}^d = r_i^{\frac{1}{\rho_\gamma - 1}} \psi_i^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} \left(r_i^{\frac{\rho_\gamma}{\rho_\gamma - 1}} \psi_i^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} + r_j^{\frac{\rho_\gamma}{\rho_\gamma - 1}} \psi_j^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} \right)^{\frac{-1}{\rho_\gamma}} (1+\gamma), \tag{10}$$

$$\gamma_{ij}^f = r_j^{\frac{1}{\rho_\gamma - 1}} \psi_i^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} \left(r_i^{\frac{\rho_\gamma}{\rho_\gamma - 1}} \psi_i^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} + r_j^{\frac{\rho_\gamma}{\rho_\gamma - 1}} \psi_j^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} \right)^{\frac{-1}{\rho_\gamma}} (1+\gamma), \tag{11}$$

where $\psi_i^{\frac{-\rho_\gamma}{\rho_\gamma - 1}} \left(\psi_j^{\frac{-\rho_\gamma}{\rho_\gamma - 1}}\right)$ is a weight attributed to domestic (foreign) rental rate for capital, and ρ_γ determines the elasticity of substitution between domestic and foreign physical capital for setting up a plant in j. For 2-country HMNEs or VMNEs with headquarters in i and a plant in j, we maintain $1+\gamma = \left[\left(\psi_i \gamma_{ij}^d\right)^{\rho_\gamma} + \left(\psi_j \gamma_{ij}^f\right)^{\rho_\gamma}\right]^{1/\rho_\gamma}$. For 3-plant HMNEs of i, $1+\gamma = \frac{1}{2}\left\{\left[\left(\psi_i \gamma_{ij}^d\right)^{\rho_\gamma} + \left(\psi_j \gamma_{ij}^f\right)^{\rho_\gamma}\right]^{1/\rho_\gamma} + \left[\left(\psi_i \gamma_{iROW}^d\right)^{\rho_\gamma} + \left(\psi_{ROW} \gamma_{iROW}^f\right)^{\rho_\gamma}\right]^{1/\rho_\gamma}\right\}$, where γ_{iROW}^d and γ_{iROW}^f are defined analogously.

Firms are assumed to maximize profits given the technologies and the demand relationships suggested above. Let c_{Xi} denote marginal production costs of differentiated good X in country i, then the profit functions consequently are:

$$\pi^{n_i} = (p_{Xi} - c_{Xi}) \sum_{j=1}^{3} x_{ij} - w_{Si} - r_i. \tag{12}$$

$$\begin{aligned}\pi^{h_{3,i}} &= \sum_{j=1}^{3} (p_{Xj} - c_{Xj}) x_{jj} - \left[1 + \sum_{j \neq i} \delta_{ij}^d + \sum_{j \neq i} \zeta_{ij}^d \mu\right] w_{Si} - \sum_{j \neq i} \delta_{ij}^f \mu w_{Sj} \\ &- \sum_{j \neq i} \zeta_{ij}^f w_{Sj} - \left[1 + \sum_{j \neq i} \gamma_{ij}^d\right] r_i - \sum_{j \neq i} \gamma_{ij}^f r_j. \end{aligned} \tag{13}$$

$$\begin{aligned}\pi^{h_{2,ij}} &= (p_{Xi} - c_{Xi}) x_{ii} + (p_{Xj} - c_{Xj}) x_{jj} - \left[1 + \delta_{ij}^d + \zeta_{ij}^d \mu\right] w_{Si} - \left[\delta_{ij}^f \mu + \zeta_{ij}^f\right] w_{Sj} \\ &- \left[1 + \gamma_{ij}^d\right] r_i - \gamma_{ij}^f r_j. \end{aligned} \tag{14}$$

$$\begin{aligned}\pi^{v_{ij}} &= (p_{Xj} - c_{Xj}) \sum_{k=1}^{3} x_{jk} - \left[1 + \delta_{ij}^d + \zeta_{ij}^d \mu\right] w_{Si} - \left[\delta_{ij}^f \mu + \zeta_{ij}^f\right] w_{Sj} \\ &- \gamma_{ij}^d r_i - \gamma_{ij}^f r_j. \end{aligned} \tag{15}$$

For i-based NEs, $w_{Si}+r_i$ denotes the fixed costs of setting up a national firm's headquarters and single plant in i. For i-based 3-plant HMNEs, the capital and skilled labor usage at home and abroad consists of the following components. First, an amount of $1 + \sum_{j \neq i} \delta_{ij}^d$ of domestic skilled workers per firm is used in the setup of the headquarters in the parent country. These workers are paid and employed in the parent country; hence, they reside and consume there. Second, an amount of $\sum_{j \neq i} \delta_{ij}^f$ skilled workers per MNE are needed from the host country to help setup the headquarters in the parent country. Hence, $\sum_{j \neq i} \delta_{ij}^f$ is the number of country i immigrants from country j generated per 3-plant HMNE to set up headquarters in the parent country. These immigrants are paid the corresponding wage in their origin country plus the premium, but they consume in the MNEs' parent country. Third, an amount of $\sum_{j \neq i} \zeta_{ij}^d$ skilled workers per firm is used from the parent country to setup and oversee foreign production sites. These workers are paid the country i wage plus the premium, but they reside and consume abroad. Accordingly, $\sum_{j \neq i} \zeta_{ij}^d$ denotes the number of skilled emigrants generated per 3-plant HMNE in i. Fourth, $\sum_{j \neq i} \zeta_{ij}^f$ is the number of foreign skilled workers employed by country i MNEs to setup foreign plants. This last category of workers is paid the wage in the respective host country, where these workers also consume. The total volume of migration generated by a 3-plant HMNE is $\sum_{j\neq i}\zeta_{ij}^{d}+\sum_{j\neq i}\delta_{ij}^{f}$. Net emigration or immigration generated by such a firm depends on whether $\sum_{j\neq i}\zeta_{ij}^{d}>\sum_{j\neq i}\delta_{ij}^{f}$ or the converse holds. The definitions of skilled labor usage and migration flows of two-plant HMNEs and VMNEs are analogous, drawing on those skilled labor markets where either headquarters or plants are actually endogenously located. Recall that δ_{ij}^{d}, δ_{ij}^{f}, ζ_{ij}^{d}, ζ_{ij}^{f}, γ_{ij}^{d}, and γ_{ij}^{f} are all endogenously determined by the MNEs' profit-maximizing decisions, taking into account relative factor prices in domestic and foreign markets.

According to those definitions, the total number of skilled workers from country i used abroad in country j (M_{ij}) is:

$$M_{ij}=\zeta_{ij}^{d}\left(h_{3,i}+h_{2,ij}+v_{ij}\right)+\delta_{ji}^{f}\left(h_{3,j}+h_{2,ji}+v_{ji}\right). \tag{16}$$

Of course a country's total emigration (i.e., the number of skilled expatriates) is then $M_i=\sum_{j\neq i}M_{ij}$. According to (16), the pattern (direction) of *aggregate* net migration, $M_i=\sum_{j\neq i}M_{ij}$, depends not only on the pattern of intra-firm net migration (which is different across types of MNEs), but also on the importance of domestic relative to foreign MNEs in general and the relative importance of particular types of MNEs.

Following Bergstrand and Egger (2007, 2008), bilateral FDI from i to j may be defined as the usage of i's physical capital in j. Similar to bilateral migration, bilateral (nominal) FDI is defined as

$$FDI_{ij}=\gamma_{ij}^{d}\left(h_{3,i}+h_{2,ij}+v_{ij}\right)r_i. \tag{17}$$

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. Two conditions characterize models in this class. First, profit maximization ensures markup pricing equations:

$$p_{Xi}=\frac{c_{Xi}(\varepsilon-1)}{\varepsilon}. \tag{18}$$

Second, free entry and exit ensure:

$$
\begin{aligned}
w_{Si} + r_i \quad &\geq \quad \frac{c_{Xi}}{-\varepsilon} \sum_{j=1}^{3} x_{ij} \quad \perp \quad (n_i) && (19) \\
\left[1 + \sum_{j \neq i} \delta_{ij}^{d} + \sum_{j \neq i} \zeta_{ij}^{d} \mu \right] w_{Si} + \sum_{j \neq i} \delta_{ij}^{f} \mu w_{Sj} + \sum_{j \neq i} \zeta_{ij}^{f} w_{Sj} \quad &+ \quad \left[1 + \sum_{j \neq i} \gamma_{ij}^{d}\right] r_i + \left[\sum_{j \neq i} \gamma_{ij}^{f} r_j\right] \\
&\geq \quad \sum_{j=1}^{3} \frac{c_{Xj}}{-\varepsilon} x_{jj} \quad \perp \quad (h_{3,i}) && (20) \\
\left[1 + \delta_{ij}^{d} + \zeta_{ij}^{d} \mu\right] w_{Si} + \left[\delta_{ij}^{f} \mu + \zeta_{ij}^{f}\right] w_{Sj} + \left[1 + \gamma_{ij}^{d}\right] r_i + \gamma_{ij}^{f} r_j \quad &\geq \quad \frac{c_{Xi} x_{ii} + c_{Xj} x_{jj}}{-\varepsilon} \perp (h_{2,ij}) && (21) \\
\left[1 + \delta_{ij}^{d} + \zeta_{ij}^{d} \mu\right] w_{Si} + \left[\delta_{ij}^{f} \mu + \zeta_{ij}^{f}\right] w_{Sj} + \gamma_{ij}^{d} r_i + \gamma_{ij}^{f} r_j \quad &\geq \quad \frac{c_{Xj}}{-\varepsilon} \sum_{j=1}^{3} x_{ji} \quad \perp \quad (v_{ij}), && (22)
\end{aligned}
$$

where $\perp$ indicates that at least one of the adjacent conditions has to hold with equality.

2.3 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 $Y_i = U_i$ ($i = 1, 2, 3$). In the presence of positive trade costs, we assume country 1 is the numéraire; hence, $p_{Y1} = w_{U1} = 1$.

2.4 Factor-Endowment Constraints and Total Factor Income

We assume that, in equilibrium, all factors are fully employed. The formal factor-endowment constraints for physical capital (K_i), skilled labor (S_i) and unskilled labor (U_i) read as follows:

$$
\begin{aligned}
K_i &\geq a_{KXi}\Big\{\Big(n_i+\sum_{j\neq i}v_{ji}\Big)\sum_{j=1}^{3}x_{ij}+x_{ii}\Big[\sum_{j=1}^{3}h_{3,j}+\sum_{j\neq i}\Big(h_{2,ij}+h_{2,ji}\Big)\Big]\Big\} \qquad (23)\\
&+\ n_i+\Big(1+\sum_{j\neq i}\gamma_{ij}^{d}\Big)h_{3,i}+\sum_{j\neq i}\Big[\Big(1+\gamma_{ij}^{d}\Big)h_{2,ij}+\gamma_{ij}^{d}v_{ij}+\gamma_{ji}^{f}\Big(h_{3,j}+h_{2,ji}+v_{ji}\Big)\Big]\\
&\perp\quad (r_i)\\
S_i &\geq a_{SXi}\Big\{\Big(n_i+\sum_{j\neq i}v_{ji}\Big)\sum_{j=1}^{3}x_{ij}+x_{ii}\Big[\sum_{j=1}^{3}h_{3,j}+\sum_{j\neq i}\Big(h_{2,ij}+h_{2,ji}\Big)\Big]\Big\} \qquad (24)\\
&+\ n_i+\Big(1+\sum_{j\neq i}\Big(\delta_{ij}^{d}+\zeta_{ij}^{d}\Big)\Big)h_{3,i}+\sum_{j\neq i}\Big[\Big(1+\Big(\delta_{ij}^{d}+\zeta_{ij}^{d}\Big)\Big)\Big(h_{2,ij}+v_{ij}\Big)\\
&+\ \Big(\delta_{ji}^{f}+\zeta_{ji}^{f}\Big)\Big(h_{3,j}+h_{2,ji}+v_{ji}\Big)\Big]\quad\perp\quad (w_{Si})\\
U_i &\geq a_{UXi}\Big\{\Big(n_i+\sum_{j\neq i}v_{ji}\Big)\sum_{j=1}^{3}x_{ij}+x_{ii}\Big[\sum_{j=1}^{3}h_{3,j}+\sum_{j\neq i}\Big(h_{2,ij}+h_{2,ji}\Big)\Big]\Big\} \qquad (25)\\
&+\ \sum_{j=1}^{3}\Big(Y_{ij}(1+\tau_{Yij})\Big)\quad\perp\quad (w_{Ui}),
\end{aligned}
$$

where $a_{\ell Xi}$ is the unit input requirement of factor $\ell=\{K,S,U\}$ in sector X and country i to produce one unit of output, and we should recall that $t_{Yij}=1$ if $i=j$.

Total factor income (GDP) in country i – i.e., expenditures by both indigenous and foreign factor owners residing there – is then determined as

$$
\begin{aligned}
E_i &= U_iw_{Ui}+S_iw_{Si}+K_ir_i+\sum_{k\neq i}\delta_{ik}^{f}w_{Sk}\mu(h_{2,ik}+h_{3,i}+v_{ik})\\
&\quad+\sum_{k\neq i}\zeta_{ki}^{d}w_{Sk}\mu(h_{2,ki}+h_{3,k}+v_{ki})-\sum_{k\neq i}\delta_{ki}^{f}w_{Si}(h_{2,ki}+h_{3,k}+v_{ki})\\
&\quad-\sum_{k\neq i}\zeta_{ik}^{d}w_{Si}(h_{2,ik}+h_{3,i}+v_{ik})+\sum_{k\neq i}\gamma_{ki}^{d}r_k(h_{2,ki}+h_{3,k}+v_{ki})\\
&\quad-\sum_{k\neq i}\gamma_{ik}^{d}r_i(h_{2,ik}+h_{3,i}+v_{ik})+\sum_{k\neq i\neq j}p_{Xk}x_{ki}(n_k+v_{ik}+v_{jk})b_{Xki}\\
&\quad+\sum_{k\neq i}p_{Yk}Y_{ki}b_{Yki}. \qquad (26)
\end{aligned}
$$

3 Calibration of the Model

The complexity of the model (including the complementarity problem in the optimization) introduces a high degree of nonlinearity so that it cannot be solved analytically. As in Markusen (2002) and Bergstrand and Egger (2007, 2008), 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 countries, developed (DC) vs. developing (LDC) economies. To limit the scope, we focus initially on bilateral flows between two developed economies, assuming the third economy (ROW) is also developed. However, later in the analysis, we consider differences in relative factor endowments. We use GAMS for our numerical analysis.

3.1 Values of (Exogenous) Factor Endowment, Trade Cost, Investment Cost, and Migration 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). We use plausible values for bilateral final goods trade according to data from the United Nations' (UN's) COMTRADE data. In particular, the mean of final goods bilateral transport cost factors [(cif - fob)/fob] among the developed countries is about 10 percent, which we use in our simulations.

Furthermore, focusing on relationships among the developed countries, we set tariffs at 2 percent, which is in line with tariff rates in Jon Haveman's TRAINS data for the 1990s and these countries. For the incremental capital usage of setting up a foreign subsidiary (γ) we use a value of 10 percent, similar to trade costs. For expatriates, we assume a premium paid on expatriates' wages of 20 percent ($\mu = 1.2$)

3.2 Utility and Technology Parameter Values

Consider first the utility function. In equation (1), 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 UN's 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 (5) 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, irrespective of the country considered.

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 a ratio that is larger than unity for both (3-plant and 2-plant) HMNEs and VMNEs. The parameters determining the composition of domestic and foreign skilled labor and capital usage are as follows: $\delta = 0.01$, $\zeta = 0.001$, $\rho_\delta = 0.9$, $\rho_\zeta = 0.9$, $\rho_\gamma = 0.5$, and $\psi_i = 0.5 \ \forall \ i \in \{1, 2, 3\}$.

4 A Theoretical Rationale for *Two-Way* Skilled Migration and for Estimating Gravity Equations

We now have a framework to motivate a theoretical rationale for two-way expatriate flows and for estimating gravity equations for two-way bilateral trade, FDI, FAS, *and* highly skilled migration flows. The two central features in a gravity equation explaining bilateral flows are the product of the two regions' GDPs and measures of bilateral "frictions" – typically, distance between economic centers, land adjacency, common language, etc. In this section, we focus on the roles of economic size and similarity, but address the roles of bilateral frictions later.

4.1 Economic Size and Similarity

To address first the expected relationships between exporter (home) and importer (host) GDPs, we focus initially on a frictionless economy. We address two key features of a frictionless gravity equation. First, in a simple theoretical world of N ($N > 2$) countries, one final differentiated good, no trade costs, but internationally *immobile* factors (labor and capital), we know from the international trade gravity-equation literature that the trade flow from country i to country j in any year ($Flow_{ij}$) will be determined by:

$$Flow_{ij} = GDP_i GDP_j / GDP^W, \tag{27}$$

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

$$lnFlow_{ij} = -lnGDP^W + lnGDP_i + lnGDP_j. \tag{28}$$

However, the standard frictionless trade gravity equation can be altered algebraically to separate influences of economic size (GDP_i+GDP_j) from similarity (s_is_j), where $s_i = GDP_i/(GDP_i + GDP_j)$ and analogously for j:

$$Flow_{ij} = GDP_iGDP_j/GDP^W = (GDP_i + GDP_j)^2(s_is_j)/GDP^W. \tag{29}$$

This formulation illustrates that $Flow_{ij}$ is a positive function of the economic size and of the economic similarity of countries i and j. When countries i and j are identical in economic size ($s_i = s_j = 1/2$), s_is_j is at a maximum. In log-linear form:

$$lnFlow_{ij} = -lnGDP^W + 2ln(GDP_i + GDP_j) + ln(s_is_j). \tag{30}$$

Second, while the gravity equation is familiar in algebraic form, it will be useful to *visualize* the frictionless "gravity" relationship. Figure 1c illustrates the frictionless gravity-equation relationship between bilateral trade flows, GDP size, and GDP similarity in equation (29) for an arbitrary hypothetical set of country GDPs ($N > 2$). We first explain the figure's axes and labeling. The lines on the y-axis in the bottom plane range from 1 to 2.2. The y-axis indexes the joint economic size of countries i and j; line 1 denotes the smallest combination of GDPs and line 2.2 denotes the largest combination. The GDP values are scaled to this index with the range tied to our *World Development Indicators* data set for GDPs, to be used shortly. The x-axis is indexed from 0 to 1. Each line represents i's share of both countries GDPs; the *center line* represents 50 percent, or *identical GDP shares for i and j.* The z-axis measures the "flow" from i to j as determined by equation (29), which is a simple algebraic transformation of typical (frictionless) gravity equation (27).

4.2 Empirically Predicted Trade, FDI, and Skilled Migration Flows

The gravity equation has a long history in the literatures on bilateral aggregate trade, FDI, and migration flows. In fact, the earliest uses of the gravity equation were to model *migration* flows, cf., Ravenstein (1885, 1889). Since then, the gravity equation has been used extensively to model migration flows, cf., Zipf (1946), Stewart (1948), Isard (1975), Sen and Smith (1995). The gravity model was first adopted for studying international trade flows in Tinbergen (1962) and Linnemann (1966), and is well established in the trade literature. Among the trade, FDI/FAS, and migration literatures, the gravity equation was adopted last in estimating FDI/FAS flows. However, as Blonigen et al. (2007) note, the "gravity model is arguably the most widely used empirical specification for FDI" (p. 1309).

In this section, we discuss the results of estimating typical gravity equations using Poisson Quasi-Maximum-Likelihood (PQML) for goods exports from country i to country j ($Trade_{ij}$), foreign affiliate sales of MNEs headquartered in country i with plants in country j (FAS_{ij}), and (highly) skilled migration by individuals born in country i that are living in country j ($Expats_{ij}$).[12] The specification we use is a standard one; we start with equation (28) and include also the logarithm of bilateral distance (measured conventionally as "great-circle" distance) and dummy variables for common land border ($Adjacency_{ij}$) and common language ($Language_{ij}$). For the LHS, we use the same panel data sets for $Trade_{ij}$ and FAS_{ij} used in Bergstrand and Egger (2008, Table 2), so that we can compare our results for migration flows with those results to address interconnections in the sensitivity analysis in section 6. While data on bilateral trade flows are available for multiple years for many countries, internationally consistent data for FAS and highly skilled workers is much more difficult to obtain. The only data set for bilateral FAS for a wide array of countries is from country profiles of UNCTAD for the years 1986-2000

[12] Following recent developments in the gravity equation literature, we use PQML estimation rather than ordinary least squares to avoid bias arising from heteroskedasticity and to enable usage of zero flows (see Santos Silva and Tenreyro, 2006).

among approximately 30 countries, many of which are members of the OECD; zero FAS values were used. The trade data set is for the same countries for 1990-2000. Details are provided in the Data Appendix.

Unlike trade and FAS data which are available for multiple years for a large group of developed and developing countries, there are no internationally consistent *panel* data sets for highly skilled migration flows. However, in 2003, the OECD launched a data collection effort in collaboration with national statistical offices to obtain statistics on the foreign-born population for each OECD country by country of birth *and* educational attainment. From this, they compiled a cross-sectional data set on the stock of emigrants born in country i living in country j with tertiary-level education (highly skilled) for year 2000. This data set was used for $Expats_{ij}$; details are provided in the Data Appendix.

Table 1 presents estimates using PQML of typical gravity equations for $Trade_{ij}$, FAS_{ij}, and $Expats_{ij}$. As discussed in the introduction, most empirical work using gravity equations examines only one of these three types of flows. Our analysis examines the determinants of all three types of flows using a common econometric specification. As typically found in empirical work, country i's and country j's GDPs are positively correlated with each of the flows, and the coefficient estimates are economically and statistically significant from zero.[13] Moreover, the coefficient estimates for GDPs are not that different economically across the three flow types. The coefficient estimates for log of bilateral distance are negative and statistically significant for trade and FAS, as typically found. However, the distance coefficient estimate is effectively zero for skilled migration; we will find later that this is a consequence of this simple specification (i.e., omitted variables bias). In the fuller specification used in section 5, we will find an economically and statistically significant negative coefficient estimate for distance with migration as well. A common language has a positive effect on all three types of flows; the coefficient estimate is statistically significantly different from zero for migration flows. Adjacency has the expected positive relationship, but is statistically insignificant.[14] Thus, the gravity equation

[13]GDP coefficient estimates less than unity are common using PQML.
[14]Adjacency dummies often are insignificant using PQML.

works very well empirically for trade, FAS, and migration of highly skilled workers. The pseudo-R^2 values for all three specifications are representative.

The qualitative (not quantitative) similarity of GDP coefficient estimates across all three flow types suggests that the gravity-equation relationships between the flow, economic size, and economic similarity are similar – but not identical – for the three flow types. To confirm this, we calculated the empirically *predicted* flows using the fitted values for each regression equation. Figures 2a-2c illustrate visually the relationship between the flow, GDP size, and GDP similarity for $Trade_{ij}$, FAS_{ij}, and $Expats_{ij}$, respectively. The explanation of the axes was described above in section 4.1. A comparison of Figures 2a-2c with Figure 1c confirms that all three flows empirically are generally represented well by a gravity equation, but not the *same* gravity equation.

4.3 Theoretically Predicted Trade, FAS, and Expat Flows

Given the large and traditionally separate empirical gravity equation literatures for trade, FDI/FAS, and migration discussed earlier, the results in Table 1 may come as no surprise to the reader. However, there is to date *no formal theoretical model* that can explain simultaneously bilateral aggregate trade, FDI, FAS, and (highly skilled) migration flows – much less one that explains simultaneously *two-way* flows.

However, our numerical GE model described earlier can motivate a theoretical rationale for explaining (two-way) trade, FDI/FAS, and skilled migration flows using a gravity equation. Figures 3a-3d illustrate visually the relationships between $Trade_{ij}$, FAS_{ij}, FDI_{ij}, and $Expats_{ij}$, respectively, with the sum of two countries' GDPs and the similarity of their GDPs. As suggested by Figures 1c and 2a-2c, Figures 3a-3d imply that our theoretical model suggests that the relationship between each type of flow and GDPs is represented well by a "gravity equation." In the context of the simulation generating these flows, all values are generated using bilateral trade, investment, and migration cost levels assumed in section 3 and all three countries have identical *relative* factor endowments.

The economic rationale for the simultaneous theoretical gravity-like relationships for all four flows is the following. Consider first the flows of goods sold either by country i's exporters or by i's horizontal MNEs (HMNEs) headquartered in i with affiliates in j; since relative factor endowments are identical in these simulations, there are no vertical MNEs (VMNEs). Not surprisingly, $Trade_{ij}$ and FAS_{ij} increase when country i's and j's GDPs (absolute factor endowments) increase. However, unlike the KC model in Markusen (2002), both trade *and* FAS are maximized when countries i and j have similar economic sizes. Even though exports (produced by national enterprises, NEs) and foreign affiliate sales (produced by HMNEs) are substitutes overall, NEs and HMNEs can *coexist* – even when countries i and j are identical in *all respects* (i.e., absolute *and* relative factor endowments); two-way trade, FDI, and FAS flows exist.

The reason is that more than one factor is used in the setup of headquarters and plants, and firm (or headquarters) setups have different relative factor intensities than plant setups.[15] In our model, firm setups require only skilled labor, but plant setups require both physical capital as well as skilled labor; consequently, firm (plant) setups are *relatively* skilled-labor (physical-capital) intensive. The intuition for Figures 3a-3f can be understood by starting at the far RHS of all six figures. When j's share of the two countries' (i's and j's) GDPs is very small, country i meets the demand in j for i's products using exports rather than FAS (because of the large size of i in the world economy), given initial levels of trade, investment, and migration costs, cf., Figures 3a and 3e. 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., Figure 3e. However, output per firm in j, production, and overall demand expands proportionately more, such that exports from i to j increase, cf., Figure 3a. As j gets larger, sales of differentiated good X by i in j are met via FAS from i to j because – at given trade, investment, and migration costs – it is more profitable for HMNEs to provide i's differentiated products to country j *and* ROW. Note that as j gets even larger – moving toward 1/2 of the two countries' GDPs – FAS_{ij} increases rapidly but $Trade_{ij}$ increases at a *diminishing* rate.

Consider now Figures 3c and 3d, expat and FDI flows from i to j, respectively. When

[15]In Markusen (2002), firm and plant setups use skilled labor only, cf., p. 80.

j is small and its demand is met by relatively large country i's exports, there is no FDI or skilled migration from i to j; *national* firms' headquarters and plant setups can use only domestic skilled labor and physical capital, respectively. However, as j's share of the two countries' GDPs increases, the relative replacement of i's NEs by i's HMNEs leads to two differences. First, the need for i to provide sales in j using foreign affiliates increases the relative need to setup plants abroad. This rise in demand for physical capital in j raises its relative price, inducing FDI from i to j (in our model, captured by a flow of physical capital from i to j, in the Mundellian tradition, cf., Mundell (1957), Jones (1967), and Helpman and Razin (1983)). Moreover, the increased demand for skilled workers in i to help setup foreign affiliates in j causes a migration of skilled workers from i to j. Second, in country i, single-plant NEs are being increasingly replaced by multi-plant HMNEs. Although an HMNE requires slightly more skilled labor to be setup than an NE does, on net the relative demand for skilled labor and its relative price in i falls owing to the increase in relative demand for physical capital to setup plants, cf., Figure 3f.[16] However, HMNEs in i have access to the pool of foreign skilled labor in j (that is prohibitively costly to NEs in i). Despite the fall in the relative price of skilled labor in i, the increase in the number of HMNEs relative to NEs in i increases the flow of skilled labor born in j to i to help setup HMNE headquarters (figure not shown for brevity). These two channels combined motivate theoretically simultaneous *two-way* highly skilled migration from i to j and j to i *and* explain the *gravity-equation* relationships in Figures 3a-3d.

Although the theoretical gravity-equation relationships are qualitatively similar, they are *quantitatively* slightly different. In particular, the expats figure attains a maximum when exporter i's share of the two countries' GDPs is smaller, whereas the trade and FDI figures attain a maximum when the two countries have the same GDPs. The economic rationale is the following. As noted above, headquarter setups require only skilled labor (either domestic, foreign, or a combination), whereas plant setups require both skilled

[16]A higher price of physical capital in i due to FDI outflows raises the relative price of multi-plant HMNE firm setups, reducing the displacement of single-plant NEs and helping secure the coexistence of trade and FAS from i to j. Also, a lower price of skilled labor in i lowers the price of HMNE and NE firm setups, also securing coexistence of both types of firms, even as i and j become identical in size.

labor and physical capital (either domestic, foreign, or a combination). Any increase in j's relative economic size will increase the relative demand for physical capital in i and its price (not shown) relative to skilled labor as single-plant NEs are replaced increasingly by multi-plant HMNEs. Hence, FDI from i to j will increase faster than expats from i to j to setup foreign affiliates in j. This tends to cause FDI from i to j to reach a maximum relative to expats from i to j when j (i) is relatively smaller (larger). Although the relative positions are influenced by choices of elasticities, the effect of the multi-plant structure in HMNEs dominates, ensuring that expats from i to j will attain a maximum relative to FDI from i to j when i is relatively smaller.[17]

5 Relative Factor Endowments and the Matching of Theory and Empirics

The empirical economic literature on migration has focused upon two key factors (beyond economic size) that tend to explain bilateral aggregate migration flows: relative wage rates (often proxied by relative per capita incomes) and the costs of migration, cf., Borjas (1994, 1999), Hatton and Williamson (2002), and Grogger and Hanson (2008). Typically, researchers in migration start with the simple framework that the decision (d) of a representative individual m (m=1,...,M) to migrate from a source country i to a destination country j is expressed as:

$$d_{mij} = w_j - w_i - cost_{ij} + z_{mij}, \tag{31}$$

where w_j (w_i) is the wage rate in country j (i), $cost_{ij}$ is the direct cost of migration from i to j, and z_{mij} represents other factors that might influence the migration decision for individual m (i.e., "compensating differentials," such as family considerations). Hatton and Williamson (2002) note that the more "recent literature" has focused on selectivity

[17]A sensitivity analysis confirms this, as the peak of FDI from i to j is near the center, whereas expats from i to j always peaks when i is smaller than j.

issues, such as skill levels. Thus, the relative wage rate (and hence the decision to migrate) may be influenced by the relative returns to skills in the two countries (w_{Sj}/w_{Uj} - w_{Si}/w_{Ui}).

In this section, we use our theoretical model to explain economic determinants of highly skilled migration flows in a broad sample of countries. First, as Markusen (2002), Blonigen (2005), and others have noted for trade, FDI, and FAS flows, the relationships between flows and their economic determinants are likely to be complex, nonlinear, and nonmonotonic. Consequently, as in Bergstrand and Egger (2008), rather than trying to specify *ex ante* a "central" regression specification for the determinants of bilateral aggregate (highly skilled) migration flows, we examine empirically a virtually exhaustive set of "theoretically agnostic" nonlinear parameter specifications and then choose *ex post* the best specification (highest pseudo-R^2). Second, we then use the numerical version of our model to theoretically predict the (predicted) migration flows from the empirical specification. We find that our model predicts the migration flows quite well; we show that the "theoretical" and "empirical" surfaces of the Edgeworth boxes for migration flows are very similar and the correlation coefficient estimates of the theoretically and empirically predicted flows are large and statistically significant. Third, in the context of our model, both migration flows and factor prices are endogenous; we cannot demonstrate that relative skilled wage rates in countries i and j *cause* skilled migration flows. However, given the GE structure of our model, we can show that our results are consistent with one of two "prominent features" of international labor movements noted recently in Grogger and Hanson (2008). We show that expats migrate from i to j when relative skilled-to-unskilled wage rates are high in j relative to that in i, an empirical finding referred to as "positive sorting."[18] Finally, our theoretical model helps to decompose *quantitatively* the amount of expat flows associated with horizontal MNEs versus vertical MNEs. We find two prominent results. First, in the context of the model, actual highly skilled migration flows (including expats) are determined *both* by horizontal *and* vertical MNEs. Second,

[18] As noted earlier, since unskilled labor is assumed immobile internationally in our model, we cannot address here the other "prominent feature" of international migration, known as "positive selection."

our results suggest that the bulk of expats flows are *two-way* flows motivated by *horizontal* MNEs.

5.1 Econometric Issues and Costs of Migration

We considered *ex ante* a wide array of different possible econometric specifications for the RHS variables suggested by our theoretical model. For instance, consider country i's and country j's shares of their joint stock of skilled labor. We started with the share of country i in the sum of i's and j's total skilled labor stock (s_i). We calculated i's share for the two countries' physical capital stock (k_i) and i's share for the two countries' unskilled labor stock (u_i). One reason for calculating these shares is that the theoretical model can be cast in terms of an "Edgeworth box," a standard tool for analyzing trade and FAS flows; the empirical country factor shares are directly related to the theoretical factor shares. We also created variables which interact (multiplicatively) these shares with each other, with the total GDP size of the two countries, and with the similarity of the two countries' GDPs. We also created variables which interact multiplicatively these country factor shares with the trade-cost and investment-cost variables of the two countries. We included logs of the shares and logs of the interactive variables just described. We included the logs of the total endowments of countries i and j of their skilled labor, unskilled labor, and physical capital stocks. Finally, we included several other control variables common to migration studies that likely influence the direct costs of migration, such as the log of bilateral distance and dummies for a common land border, common official language, common spoken language, previous colonial relationship, post-World-War-II colonial relationship, and common colonizing country.

As discussed in the previous section, our OECD data set is constrained to a cross-section of country pairs in year 2000; consequently, any influence of the *ROW* is captured in the constant. We ran Poisson Quasi-Maximum-Likelihood (PQML) estimates of every possible specification variant using the set of potential determinants. The reason for the PQML estimation method is the following. First, because of the likely multiplicative relationship between levels of flows and their economic determinants and Jensen's inequality, in the presence of heteroskedasticity OLS estimation of log-linearized equations may lead to biased coefficient estimates for RHS variables. The PQML estimation method can address this concern, cf., Santos Silva and Tenreyro (2006). Second, PQML can be applied with dependent variables that include zero and positively-valued observations, such as migration, trade, and FAS flows. PQML exploits variation from both zero and non-zero observations, and our theoretical model predicts a large number of zeros and our cross-section of empirical migration flows has a large number of zeros.

The PQML specification that generated the highest pseudo-R^2 for migration flows is listed in Table 2. Column (2) of Table 2 presents the coefficient estimates for the migration specification and column (3) presents the z-statistics for those coefficient estimates. The specification is very similar to that in Bergstrand and Egger (2008) for the empirical prediction of bilateral trade and FAS flows; this was intentional so as to demonstrate that a *common* set of fundamental economic variables can explain simultaneously bilateral trade, FAS, and highly skilled migration flows. The pseudo-R^2 for the migration specification is 0.91, which exceeds the explanatory power of any previous migration study we have come across (even one including country-pair fixed effects). Thus, while we use a complex, nonlinear specification, it predicts very well highly skilled migration flows (using the "best" world bilateral highly skilled migration-flow data set presently known). The pseudo-R^2 of 0.91 for migration flows compares favorably with that for bilateral trade flows, and is slightly higher than that for bilateral FAS (0.79) found in Bergstrand and Egger (2008, Table 2).

Explaining economically the relationships between the (predicted) empirical migration flows and *all* the explanatory variables in Table 2 would require an entire book. Consequently, we focus only upon the key factors (beyond gravity) that influence the migration flows that the migration literature has emphasized: relative factor endowments (which then influence relative factor prices) and direct costs of migration. As discussed earlier, the focus of traditional and recent analysis of migration determinants has been on relative prices of labor between origin and destination economies, cf., Borjas (1994), Hatton and

Williamson (2002), and Hanson (2006). In our framework, relative prices are endogenous, influenced by (exogenous) relative factor endowments. Section 5.2 below will address the role of relative factor endowments.

However, we first address the direct costs of migration. Recent empirical analyses of migration flows have included the log of bilateral distance and dummies for a common land border, common language, common colonizing country, etc. as RHS variables. Thus, the vast bulk of studies we have come across include (log of) bilateral distance and various (country-pair) dummies. First, we discuss a set of control variables described above. Unlike in Table 1 earlier, once we control for the large set of variables mentioned above, bilateral distance now has an economically and statistically significant negative effect on the migration flow (from i to j). Furthermore, the coefficient estimates for dummies for a common land border, common official language, and common spoken language are all positively signed (as expected), economically significant, and statistically significant. The coefficient estimates for dummies for common colonizing country, common colonial relationship since World War II, and same country all were positively signed (as expected) and were economically (but not statistically) significant.

Second, since the decision to migrate in our context involves the profit-maximizing decision of MNEs, it makes sense to also include *alternative* "costs" to migration that influence firms. In our model, trade costs and investment costs alter the *relative* costs of skilled migration.[19] Consequently, our model includes these, but in complex, nonlinear ways. Consequently, we can only provide estimates of the *marginal* effects of trade costs and investment costs conditioned upon, say, the means of all the other variables. These results are available from the authors on request.

5.2 Relative Factor Endowments and Skilled Migration Flows

We now turn to the role of relative factor endowments for explaining (highly) skilled migration flows. We took the (predicted) empirical bilateral migration flows from the

[19]Unlike trade costs and investment costs, for which there are some panels of estimates, there are no cross-section, much less panel, estimates of direct migration "costs" (to our knowledge).

PQML econometric equation with the best fit and plotted them against pairs of factor endowment shares. To start with, Figure 4a plots these empirically-predicted migration flows against country i's share of the total (highly) skilled labor stock of countries i and j (denoted s_i) and against country i's share of the total unskilled labor stock of i and j (u_i). However, since there are *three* factors of production, the relationships are likely sensitive to country i's share of the total physical capital stock of i and j (k_i). For instance, Figure 4a plots these flows against s_i and u_i at the mean of k_i (0.5). However, the results are sensitive to the level of k_i; we will address later the sensitivity of these results to variation in k_i from 0 to 1. Two features of Figure 4a are worth noting. First, at any level of u_i, skilled migration increases monotonically with country i's share of the two countries' (i's and j's) skilled labor stock, s_i. Second, the increase is largest when i has a large share of the two countries' unskilled labor stock.

Theoretical Figures 4b-4e help to interpret these empirical results. Figure 4b plots the predicted *theoretical* skilled migration flows, based upon the numerical version of our model, using the same axes for ease of comparison and also at $k_i = 0.5$. First, the theoretical migration flows also tend to increase with country i's share of the two countries' skilled labor force. Second, the increase *also tends* to be largest when i's share of the two countries' unskilled labor stock is large. However, in theoretical Figure 4b, the relationship between $Expats_{ij}$ and u_i at high levels of s_i is not monotonic; the bimodal distribution for skilled migration from i to j suggests two important sources of the flows.

Figures 4c and 4d help reveal the two sources of these peaks. First, Figure 4c shows the numbers of 3-country HMNEs with headquarters based in country j with plants in countries i and ROW. When country i has a large share of the two countries' skilled (and unskilled) labor but only one-half of their physical capital, the relative abundance of skilled workers in i tends to raise the demand by horizontal MNEs in j for these skilled workers to be transferred to setup headquarters in country j, raising the expat flow from i to j.

Figure 4d shows the numbers of VMNEs with headquarters based in country i and a plant in j. When country i is abundant in skilled labor but scarce in unskilled labor (relative to j), i has a comparative advantage in providing headquarter services and "outsourcing" production of differentiated (final good) X to j, which is relatively unskilled labor abundant. However, setting up production facilities in j requires skilled workers to migrate from i to j, which explains the other peak when s_i (u_i) is high (low).

The differing sources of skilled migration (HMNEs vs. VMNEs) also helps to explain the "positive sorting" feature of the international migration literature, noted in Grogger and Hanson (2008). In the first case, skilled migration from i to j is related to relative economic *sizes* of i and j, and not relative factor endowments; this migration is tied to horizontal MNEs and so is *not* related to positive sorting. However, in the second case, skilled migration from i to j is related to relative factor endowment differences (VMNEs), in particular, a lower ratio of skilled labor relative to unskilled labor in j relative to that in i. In this case, the relative price of skilled-to-unskilled labor should be higher in j. Theoretical Figure 4e confirms the relatively higher price of skilled-to-unskilled labor in j relative to i when VMNEs based in i are relatively prevalent, which is consistent with empirical evidence of the positive correlation between skilled migration from i to j and a higher relative price of skilled-to-unskilled labor in destination countries (i.e., positive sorting and the brain drain).

An important point worth noting is that these relationships – both theoretical *and* empirical – are sensitive to the level of *physical capital*. Figure 4f confirms this empirically, showing the empirically-predicted skilled migration flows from i to j at k_i=0.7. The notable difference with Figure 4a is that there are more $Expats_{ij}$ due to a larger role for relative factor endowment differences. The corresponding theoretical figure is consistent with this, showing that $Expats_{ij}$ are associated with VMNEs even more so when $k_i = 0.7$ (not shown for brevity).[20]

A much fuller representation of the sensitivity of the empirical and theoretical skilled migration from i to j to *all levels* of physical capital shares is captured by Figures 5a-5f. For instance, Figure 5a presents the empirically-predicted skilled migration flows from i to

[20]Conversely, in the empirical surface at k_i=0.3, there is much less predicted $Expats_{ij}$ due to relative factor endowment differences, and this is confirmed in the corresponding theoretical Edgeworth box at k_i=0.3. Figures are omitted for space constraints, but are available on request.

j in relation to country i's and j's shares of the two countries' total skilled labor stocks (s_i and s_j, respectively) and total physical capital stocks (k_i and k_j, respectively) at u_i=0.3. First, as empirical Figure 5a shows, skilled migration from i to j has a peak when i is very abundant in skilled labor, but relatively scarce in physical capital and unskilled labor relative to j (k_i and u_i are both about 0.3). Second, but subtler, skilled migration reaches another peak (though smaller) when i is very *scarce* in skilled labor, but also relatively scarce in physical capital and unskilled labor (k_i and u_i are both about 0.3).

Our model can explain these two empirical results. First, Figure 5b shows the theoretical relationship between $Expats_{ij}$, s_i, and k_i. The figure captures the overall relationships suggested in empirical Figure 5a. Note that migration has its most prominent peak when i is very abundant in skilled labor, but physical capital and unskilled labor are relatively scarce. There is also another (somewhat smaller peak) when s_i is lower, and physical capital and unskilled labor in i are scarce (relative to j). Second, Figure 5c illustrates the number of 3-country HMNEs with headquarters in country i and plants in i, j, and ROW. Since country i is abundant in skilled labor (relative to j), but scarce in physical capital and unskilled labor, i can most profitably serve all three markets using HMNEs headquartered in i. With mobile skilled workers, the profitable setup of plants in j (and ROW) will use skilled workers from i (which are abundant) to help setup plants abroad, fostering $Expats_{ij}$. Third, Figure 5a implies that skilled migrants from i to j will also occur when i is scarce in skilled labor relative to j, as well as unskilled labor and physical capital. In this case, i's market can be most profitably served by HMNEs headquartered in relatively larger country j with plants also in i and ROW. Figure 5d shows that HMNEs headquartered in j will need to send skilled workers from i to j to setup j's HMNEs' headquarters.

Finally, Figure 5e shows that a large number of VMNEs headquartered in i with plants in j will transfer skilled workers from i to j to setup plants when i is extremely abundant in skilled labor, but scarce in unskilled labor and physical capital, as expected. If our model is consistent with "positive sorting" in migration, then our model should suggest that relative skilled-to-unskilled wage rates should be higher when migration flows are higher, i.e., when j is relatively scarce in skilled labor (relative to unskilled labor) relative to i. Figure 5f shows the relationship between relative skilled-to-unskilled wage rates in country j relative to country i. In the k-s Edgeworth box, skilled migration tends largely to increase as i's share of skilled labor increases, given u_i=0.3 and low levels of k_i. If "positive sorting" is present, one should find that wage rates of skilled-to-unskilled workers in j relative to i should also increase. Figure 5f confirms that they do.

In summary, we note three conclusions. First, the theoretical model tends to explain well qualitatively the empirically-predicted (highly) skilled migration flows. In fact, the correlation coefficient between the empirically predicted $Expats_{ij}$ and the theoretically predicted $Expats_{ij}$ in Figures 4a and 4b, respectively, is 0.234, which is statistically significant (with a p-value of 0.000002). The correlation coefficient between the empirically predicted $Expats_{ij}$ and the theoretically predicted $Expats_{ij}$ in Figures 5a and 5b, respectively, is 0.597, which is also statistically significant (with a p-value of 0.000001). Second, regardless of axes used, the theory suggests a common conclusion: the highly skilled migration flows can be explained by both HMNEs and VMNEs. Third, even though both types of MNEs matter, it is *two-way* skilled migration motivated by HMNEs that explains the largest predicted volumes of expat flows.

6 Interconnectedness: Explaining Bilateral Trade, FAS, *and* Skilled Migration Simultaneously

One of the important elements of our GE approach is that it allows simultaneous predictions of bilateral trade, FDI, FAS, and skilled migration flows and – in the terminology of Blonigen (2005) – *interconnectedness*. Consequently, a ready check of the robustness of the results above is to determine if the theoretical model *also* predicts bilateral trade and FAS flows. In the interest of constraining this paper to a reasonable length, we draw upon the (predicted) empirical surfaces for bilateral trade and FAS flows in Bergstrand and Egger (2008) to evaluate the robustness of our theoretical model in this paper.

6.1 Interconnectedness with FAS

We examine the predictions of our model for FAS first. Several conclusions are worth noting. First, Figure 6a illustrates the relationships between the (predicted) empirical FAS of country i's MNEs in country j, country i's share of the two countries' skilled labor force (s_i), and country i's share of the two countries' unskilled labor force (u_i) at the mean of k_i, using the PQML regression results from Bergstrand and Egger (2008, Table 2). As discussed there, FAS of i in j is at a maximum when country i is skill abundant relative to j, and when i is slightly smaller than j. Second, FAS is prominent when i is relatively scarce in skilled and unskilled labor, but as s_i increases at values of u_i the relationship between s_i and FAS of i in j is *not* monotonic.

Second, Figure 6b shows the theoretical relationship between FAS_{ij}, s_i, and u_i, also at the mean of k_i, implied by our present model. The theoretical model predicts the empirically-predicted FAS_{ij} very well, in particular the major peak. The correlation coefficient between the empirically predicted FAS_{ij} and the theoretically predicted FAS_{ij} is 0.36 (p-value = 0.0001). If country i is skilled-labor abundant (relative to unskilled labor), country i will have a comparative advantage in setting up VMNEs (rather than exporters), tending to increase FAS_{ij}, as confirmed in theoretical Figure 6c showing the number of VMNEs with headquarters in i and plants in j.

Third, Figures 6a-6c show that there is a small bias toward FAS_{ij} being higher the relatively smaller in economic size is country i relative to j. For given relative factor endowments, economically smaller countries cannot cover as many fixed costs of setups (of either plants or firms), tending to increase their comparative advantage in setting up VMNE headquarters, especially when unskilled labor is scarce, a point addressed in Markusen's knowledge-capital model.

Fourth, although the theoretical surface in Figure 6b appears to indicate that there is a monotonic relationship between the skilled-to-unskilled-labor ratio in country i relative to country j and levels of FAS_{ij}, empirical Figure 6a suggests a *nonmonotonic* relationship at very low levels of u_i. While the empirically-predicted FAS_{ij} flows cannot be decomposed into VMNEs and HMNEs, the theoretical FAS_{ij} flows can. Theoretical Figures 6c and 6d show the numbers of VMNEs with headquarters in i and plants in j and HMNEs with headquarters in i and plants in j and ROW, respectively.[21] As quite visible, VMNEs largely explain the value of FAS_{ij} in Figure 6b and (by implication) Figure 6a, although HMNEs also explain some of the FAS.

Fifth, for a robustness check of the FAS results, Figure 6e illustrates the relationship between the empirically predicted FAS_{ij}, country i's share of the two countries' physical capital stock (k_i), and country i's share of the two countries' unskilled labor force (u_i) at the mean of s_i. This figure shows that FAS_{ij} is maximized also when country i is *physical capital* abundant (relative to unskilled labor) *relative to* country j. However, an interesting result that emerges from the theoretical model is that HMNEs – not VMNEs – explain the large values of FAS_{ij} in empirical Figure 6e. Figure 6f illustrates the relationship between the number of HMNEs based in country i with plants in i, j, and ROW; the correlation coefficient between the two is 0.47 (p-value = 0.00001). As shown, for any given level of k_i, the number of HMNEs based in i increases as i becomes less unskilled-labor abundant (and smaller). As i becomes both smaller and relatively physical-capital abundant (with s_i=0.5), HMNEs become a more profitable way for i to serve the domestic and foreign markets relative to national firms, which increases both horizontal FDI by i *and* two-way skilled migration between i and j. Comparable results hold at various levels of k_i and s_i at the mean of u_i (not shown due to space constraints, but available on request).

6.2 Interconnectedness with Trade Flows

Finally, we examine the predictions of our model for $Trade_{ij}$. Several conclusions are worth noting. First, Figure 7a illustrates the relationships between the empirically predicted exports of country i to country j, country i's share of the two countries' skilled labor force (s_i), and country i's share of the two countries' unskilled labor force (u_i), at

[21]Recall that all data in the z-axis is indexed from 0 to 100 so that a direct comparison of Figures 6c and 6d can be misleading.

the mean of k_i, using the PQML regression results from Bergstrand and Egger (2008, Table 2). Three features of the empirical surface are worthy of explanation. First, the value of the trade flow from i to j ($Trade_{ij}$) tends to increase as country i's share of the two countries' skilled labor stock increases, but not monotonically. Second, the trade flow tends to be maximized when the two countries have *identical* shares of their joint unskilled labor force but i has *all* of the two countries' skilled labor. Third, and more subtle, there tends to be a notable amount of exports from i to j when country i is relative abundant in unskilled labor, but scarce in skilled labor (at the mean of k_i).

Second, theoretical Figure 7b shows the trade flow of all national exporting firms (NEs) from i to j from our model. There is a remarkable qualitative similarity of empirical Figure 7a and theoretical Figure 7b. Figure 7b captures the nonmonotonic increase of $Trade_{ij}$ as i's share of the two countries' skilled labor increases. The figure also captures that the trade flow tends to be maximized when the two countries have identical shares of their joint unskilled labor force but i has all the the two countries' skilled labor. (The third, more subtle feature of Figure 7a will be discussed shortly.) The reason for the first feature is that, for given shares of unskilled labor and physical capital, the larger is i's share of skilled labor, the lower the relative price of skilled labor in i, allowing i to serve markets j (and ROW) more profitably with single-plant NEs, given trade and investment barriers and the availability of unskilled labor and physical capital in i.

To understand the reason for the second feature that NEs are most profitable when i has half of the unskilled labor and physical capital, we draw on Figure 7c. Figure 7c presents the numbers of NEs operating in country i. This figure indicates that the *number* of i's exporters is maximized when i has all of both countries' skilled and unskilled labor, because the abundance of unskilled labor in i makes production costs low and the abundance of skilled labor (relative to physical capital) makes single-plant NEs more affordable. Now consider what happens when i's share of the two countries' unskilled labor (u_i) falls. The number of exporters and varieties produced in i decline. However, output per firm in i, production, and overall demand expands proportionately more (figures not shown, for brevity, but available), such that exports from i to j increase.

Third, the subtle feature of empirical Figure 7a is the apparent pronounced observance of trade when country i's share of unskilled labor is quite high, but i's share of skilled labor is low. In this case, the relative scarcity in i of skilled labor to create headquarters, but abundance of unskilled labor for production, makes i a perfect candidate for hosting production for VMNEs based in countries j and ROW, i.e., an export-platform country. This is confirmed theoretically by Figure 7d, which illustrates the preponderance of VMNEs headquartered in j with plants in (and potentially exports from) i; this explains the third, subtle feature of empirical Figure 7a.

Fourth, for robustness, we present briefly confirmation that the empirical and theoretical trade flow results are insensitive to the choice of axes. Figure 7e shows the relationship between empirically-predicted trade flows from i to j, i's share of the two countries' unskilled labor (u_i), and i's share of the two countries' physical capital (k_i). Figure 7f confirms convincingly the richness of the theoretical model, illustrating the relationship between NEs exports from i to j and relative factor endowments.

7 Conclusions

We have presented a formal GE framework for understanding the determinants of two-way international flows of expatriates, FDI/FAS, and trade. The general equilibrium theoretical model explains the existence of *two-way* expatriate flows, alongside explaining traditional (one-way) flows based upon relative factor endowments, while explaining simultaneously intra- and inter-industry trade and horizontal and vertical foreign affiliate sales. The numerical version of the model provides a rationale for estimating simultaneously gravity equations of aggregate bilateral trade, FDI/FAS, and highly skilled migration flows. Finally, our model is fully consistent with the prominent international migration literature result of the positive empirical correlation between bilateral skilled migration flows and the differential in relative returns to "skills" found in the migration literature (i.e., "positive sorting" and the "brain drain"). It is important to emphasize that two-way migration flows of highly skilled workers arise in the presence of horizontal *and* vertical MNEs. Both types of MNEs surface theoretically in equilibrium, and both types help to explain the *actual* pattern of bilateral FAS.

An important implication of the paper is the motivation of a formal theory of intra-industry *and intra-firm* two-way skilled migration. As is typical in the intra-industry trade and FDI literature, bilateral trade and FDI will be at a *maximum* between a pair of countries when the two are identical in (absolute) economic size (all other factors, including relative factor endowments, equal). In the spirit of the intra-industry trade-FDI literature, Figure 8a shows that two-way intra-industry *and intra-firm skilled migration* of horizontal MNEs in a pair of countries will also be at a maximum when the countries are identical in all respects. Note furthermore, as expected, that Figure 8b shows that two-way intra-firm skilled migration among vertical MNEs will be *zero* when the country pair is identical. Not surprisingly, as Figure 8c shows, overall bilateral intra-firm skilled migration is at a maximum when the two countries are identical in every economic aspect. Our MNE-based motivation for such migration, moreover, is fully consistent with the stylized fact noted earlier that – between the two largest economies in the world, the EU and the US (with approximately identical relative factor endowments) – *81 percent* of skilled migration is of *executives and managers*.

For space constraints, we have provided only a limited number of the empirical and theoretical surfaces generated by our empirical and theoretical models, respectively. Further sensitivity analysis confirms the robustness of our analysis; other figures are available on request. However, the results provided suggest a consistent result theoretically and empirically that highly skilled migration flows are influenced both by horizontal and vertical MNEs.

Finally, our analysis suggests future directions to proceed. First, the model includes only *final* goods; we have purposely avoided introducing intermediate goods – so called, "outsourcing of intermediates" – because modeling of that feature as well introduces considerably more complexity. However, future research should address this. Second, we have avoided intentionally assuming "heterogeneous productivities" of NEs and MNEs, which introduces another degree of complexity well beyond our already complex model.

Future research should address this important research as well.

Data Appendix

Bilateral migration of expatriates data are from the OECD's (2005) database *Emigration Rates for Highly Educated Persons by Country of Birth*. Only persons with a tertiary level of education are defined as highly educated. There are two different sets of expatriate data available. We use the one based on the approach of Cohen and Soto (2001), covering 95 economies. While the original data only provide emigration rates, we multiply these rates by the number of highly educated persons in the country of origin to obtain a cross-section of bilateral skilled migration flows for year 2000.

Bilateral export flows in U.S. dollars are taken from U.N.'s World Trade database and cover 36 countries and the years 1990-2000, as in Bergstrand and Egger (2008). Nominal export data in U.S. dollars have been deflated using producer prices indices (base year 2000) of the exporter country. Data on bilateral foreign affiliate sales (FAS) and stocks of foreign direct investment (FDI) are available from UNCTAD's Foreign Direct Investment On-line. We use data for the same 36 countries and years as for the export flows as in Bergstrand and Egger (2008). Similar to Carr, Markusen, and Maskus (2001), we deflated nominal foreign affiliate sales in U.S. dollars by host country producer price indices.

Physical capital stocks are computed by using the perpetual inventory method, using gross fixed capital formation and investment deflator data from the *World Development Indicators* and assuming a depreciation rate of 13.3 percent. Data on human capital endowments were kindly provided by Scott Baier and are based on information in the *World Development Indicators* on school enrollments, cf., Baier, Dwyer, and Tamura (2006). Bilateral distance was computed using "great circle" distances. The country trade and investment resistance indexes are from Carr, Markusen, and Maskus (2001), and kindly provided by Keith Maskus. We use data on GDPs from the World Bank's *World Development Indicators*). Indicator variables for common border, common official language, common spoken language, colonial relatioships ever, and colonial relationships since 1945 were made available by CEPII.

References

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

Baier, Scott L., Gerald Dwyer, and Robert Tamura (2006), How important are capital and total factor productivity for economic growth?, *Economic Inquiry* **44**(1), 23-49.

Baldwin, Richard E., Rikard Forslid, Philippe Martin, Gianmarco Ottaviano, and Frederic Robert-Nicoud (2003), *Economic Geography and Public Policy*, Princeton University Press, Princeton, New Jersey.

Bergstrand, Jeffrey H. and Peter Egger (2007), A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises, *Journal of International Economics* **73**(2), 278-308.

Bergstrand, Jeffrey H. and Peter Egger (2008), Finding vertical multinational enterprises, working paper.

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

Blonigen, Bruce A., Ronald B. Davies, Glen R. Waddell and Helen T. Naughton (2007), FDI in space: Spatial autoregressive relationships in foreign direct investment, *European Economic Review* **51**(5), 1303-1325.

Borjas, George J. (1994), The economics of immigration, *Journal of Economic Literature* **32**(4), 1667-1717.

Borjas, George J. (1999), The economic analysis of immigration, in Orley C. Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, North-Holland, Amsterdam.

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

Cohen, D. and M. Soto (2001), Growth and human capital: (Good data, good results, OECD Development Centre Working Paper No. 179 (http://www.oecd.org/dataoecd/33/13/2669521.xls).

Dixit, Avinash K. and Joseph E. Stiglitz (1977), Monopolistic competition and optimum product diversity, *American Economic Review* **67**(3), 297-308.

Docquier, Frederic and Abdeslam Marfouk (2004), Measuring the international mobility of skilled workers (1990-2000), World Bank Working Paper.

Ekholm, Karolina, Rikard Forslid and James R. Markusen (2007), Export-platform foreign direct investment, *Journal of the European Economic Association* **5**(4), 776-795.

Fujita, M., Paul Krugman, and Anthony J. Venables (1999), *The spatial economy: Cities, regions and international trade*, MIT Press, Cambridge, Massachusetts.

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

Griliches, Zvi (1969), Capital-skill complementarity, *Review of Economics and Statistics* **51**(4), 465-468.

Grogger, Jeffrey and Gordon H. Hanson (2008), Income maximization and the selection and sorting of international migrants, *NBER* Working Paper 13821.

Hanson, Gordon H. (2006), Illegal migration from Mexico to the United States, *Journal of Economic Literature* **44**(4), 869-924.

Hatton, Timothy J. and Jeffrey G. Williamson (2002), What fundamentals drive world migration?, *NBER* Working Paper 9159.

Hatton, Timothy J. and Jeffrey G. Williamson (2005), *Global Migration and the World Economy*, MIT Press, Cambridge, Massachusetts.

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

Head, Keith and John Ries (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.

Helpman, Elhanan (2006), Trade, FDI, and the organization of firms, *Journal of Economic Literature* **94**(3), 589-630.

Helpman, Elhanan and Assaf Razin (1983), Increasing returns, monopolistic competition, and factor movements: A welfare analysis, *Journal of International Economics* **14**(3/4), 263-276.

Isard, Walter (1975), A simple rationale for gravity-model-type behavior, *Papers of the Regional Science Association* **35**(1), 25-30.

Jones, Ronald W. (1967), International capital movements and the theory of tariffs and trade, *Quarterly Journal of Economics* **81**(1), 1-38.

Linnemann, Hans (1966), *An Econometric Study of International Trade Flows*, North-Holland Publishing, Amsterdam.

Markusen, James R. (2002), *Multinational Firms and the Theory of International Trade*, MIT Press, Cambridge, Massachusetts.

Mundell, Robert A. (1957), International trade and factor mobility, *American Economic Review* **47**(3), 321-335.

Organization for Economic Cooperation and Development (2005), *Emigration Rates for Highly Educated Persons by Country of Birth*, Paris.

Ravenstein, Ernest George (1885), The laws of migration: Part 1, *Journal of the Statistical Society of London* **48**(2), 167-235.

Ravenstein, Ernest George (1889), The laws of migration: Part 2, *Journal of the Royal Statistical Society* **52**(2), 241-305.

Santos Silva, João M.C. and Silvana Tenreyro (2006), The log of gravity, *Review of Economics and Statistics* **88**(4), 641-658.

Slaughter, Matthew J. (2000), Production transfer within multinational enterprises and American wages, *Journal of International Economics* **50**(2), 449-472.

Sen, Ashish and Tony E. Smith (1995), *Gravity Models of Spatial Interaction Behavior* Springer-Verlag, Berlin.

Stewart, John Q. (1948), Demographic Gravitation: Evidence and Application, *Sociometry* **11**(1/2), 31-58.

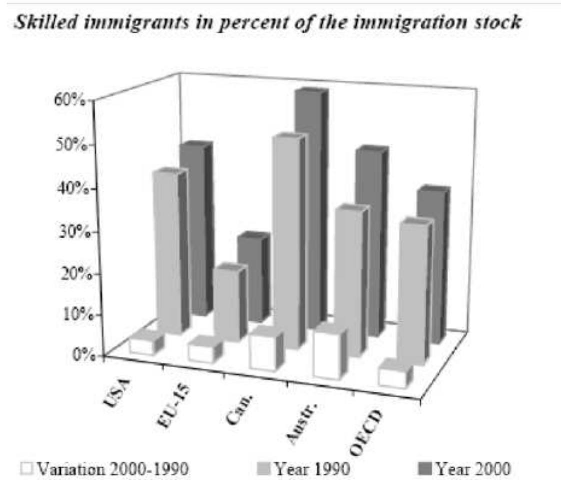
Straubhaar, Thomas and Achim Wolter (1997), Globalisation, Internal Labour Markets and the Migration of the Highly Skilled, *Intereconomics* **32**(4), 174-180.

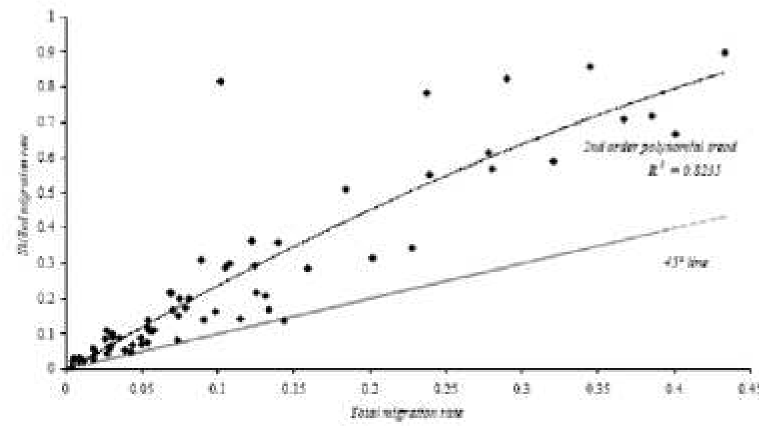
Tinbergen, Jan (1962), *Shaping the World Economy: Suggestions for an International Economic Policy*, The Twentieth Century Fund, New York.

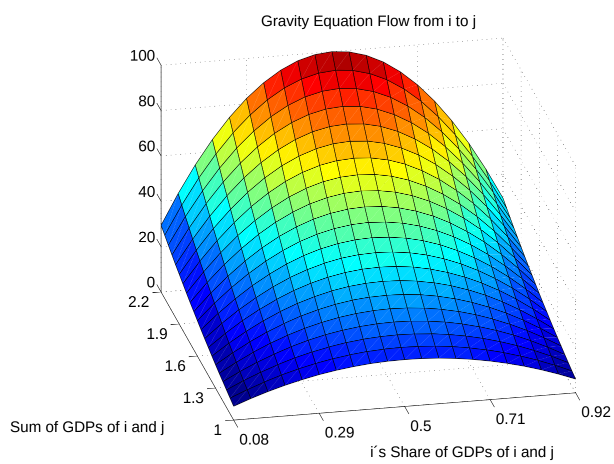
Yeatman, Perry and Stacie Nevadomski Berden (2007), *Getting Ahead by Going Abroad* HarperCollins, New York, NY.

Zipf, George K., (1946), The P_1 P_2 / D hypothesis: On the intercity movement of persons, *American Sociological Review* **11**(6), 677-686.

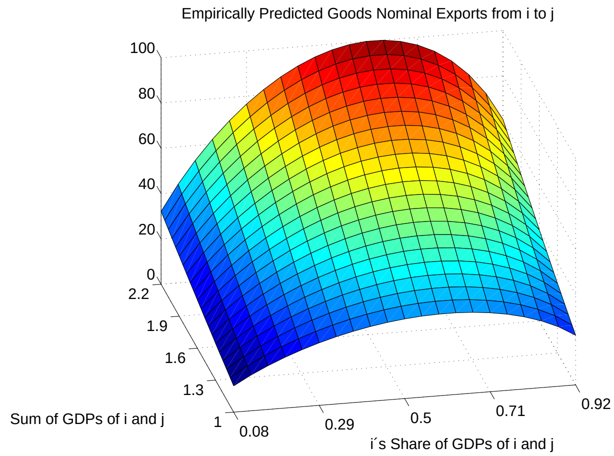
Figures 1a, 1b, 1c

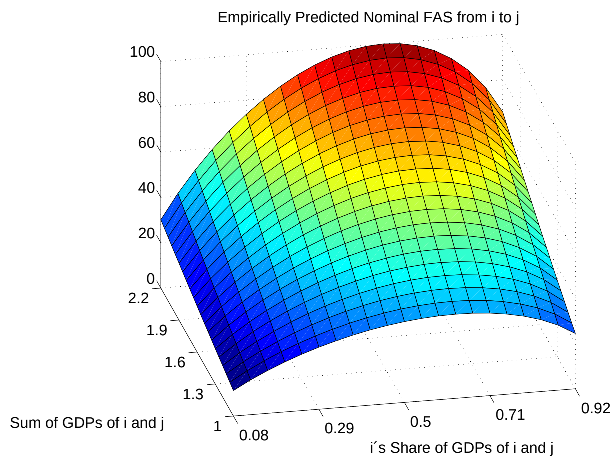


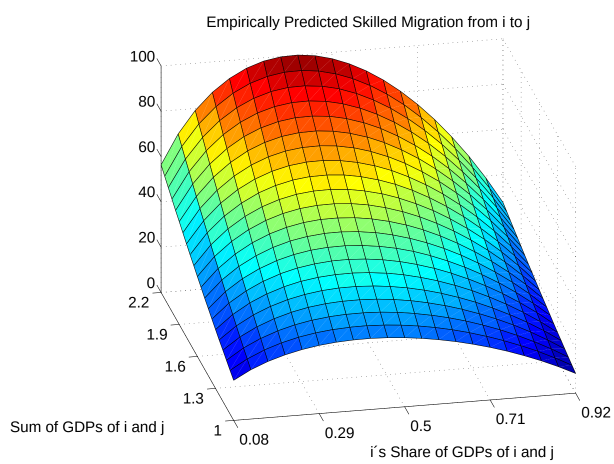




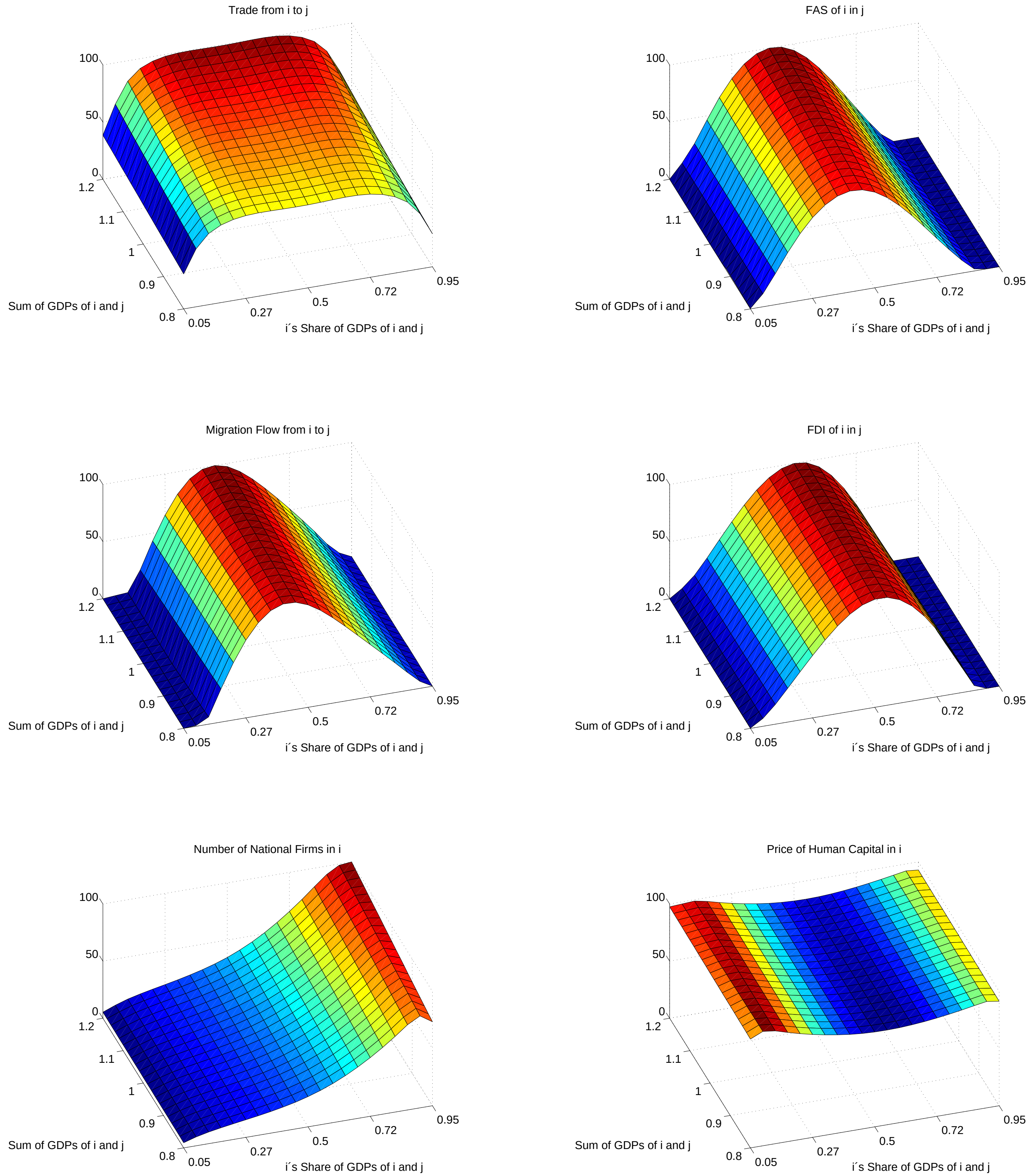
Figures 2a, 2b, 2c



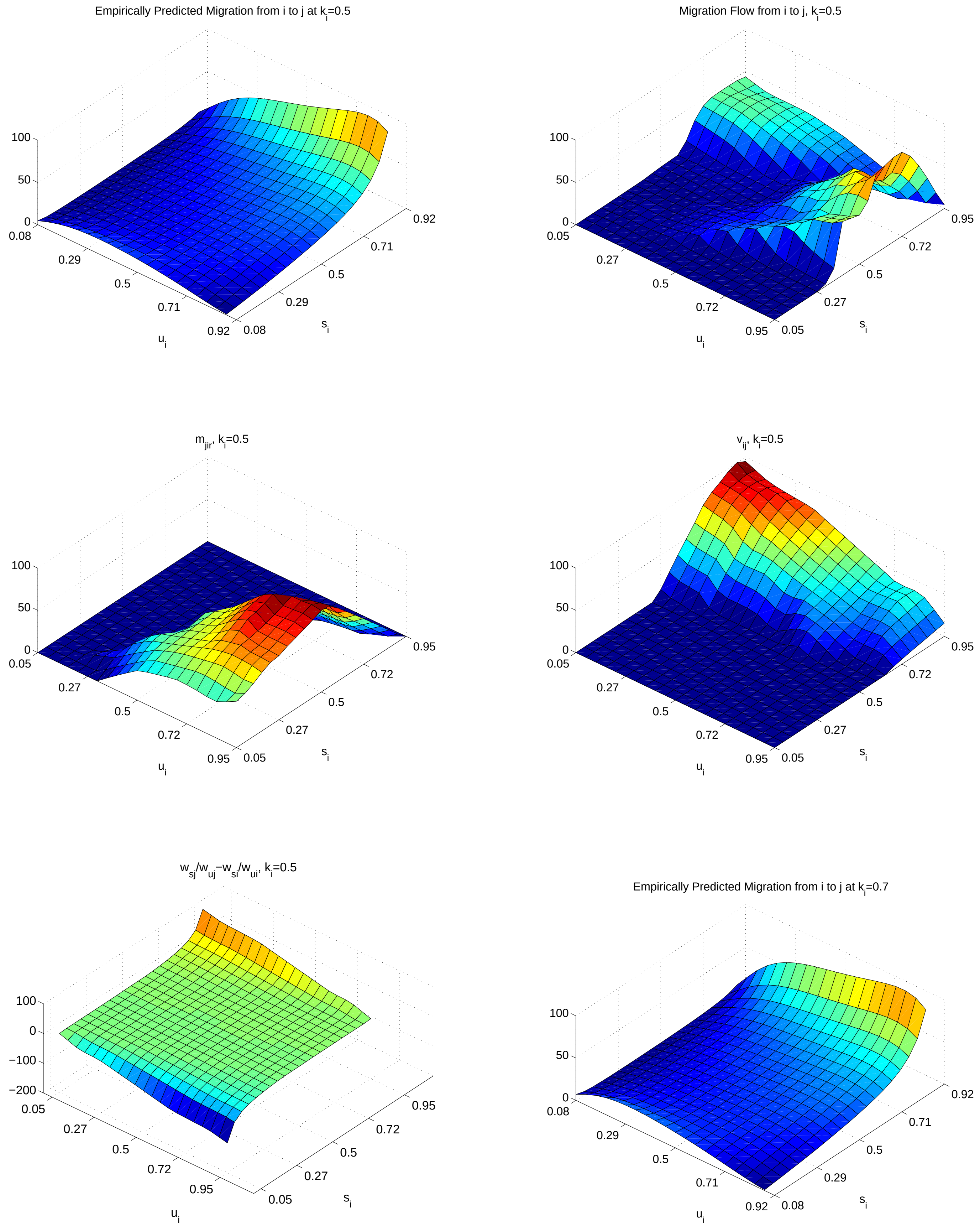




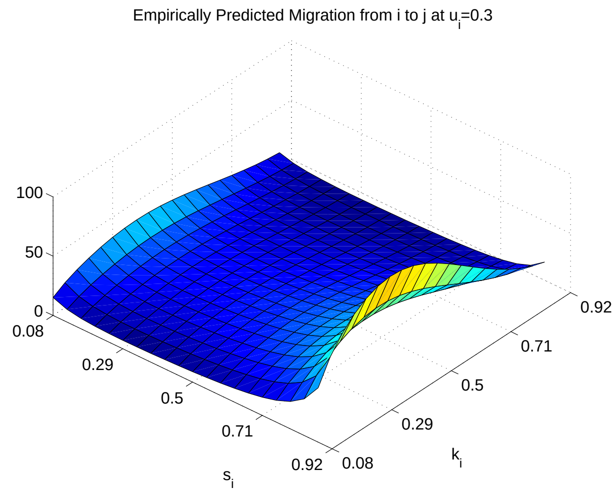
Figures 3a-3f

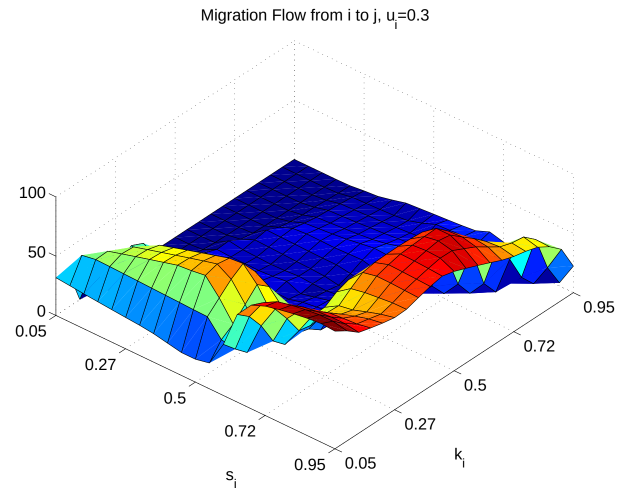


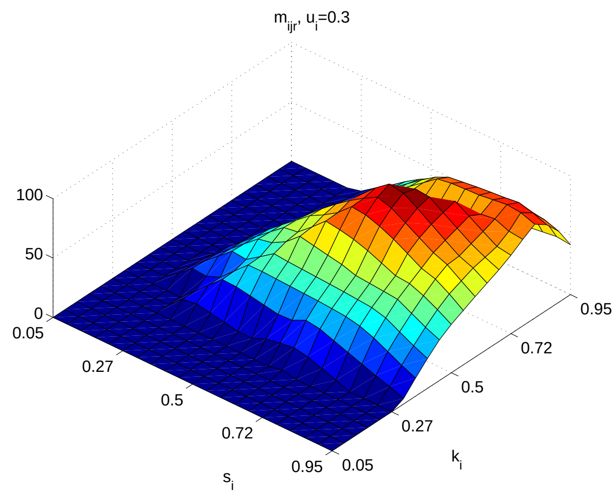
Figures 4a-4f

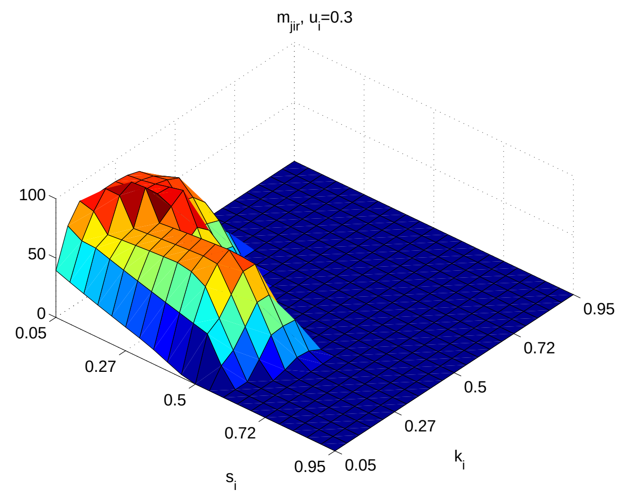


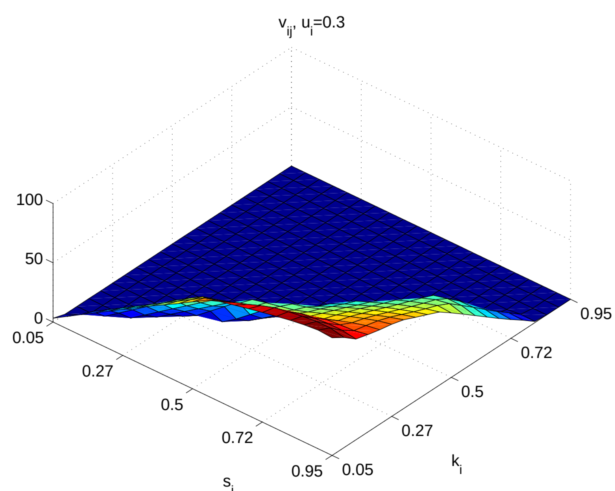
Figures 5a-5f

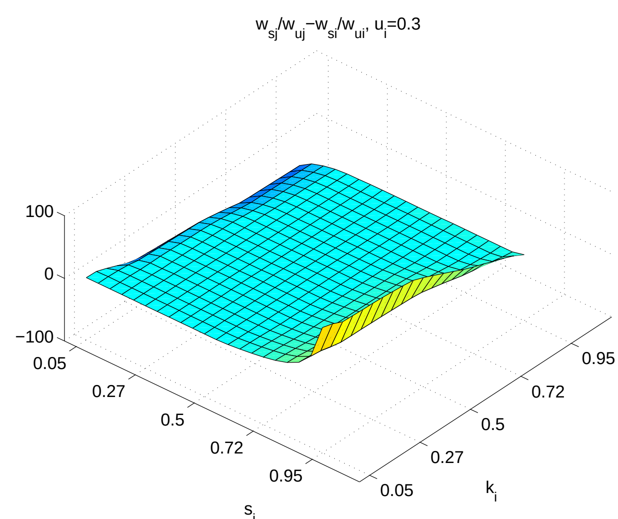




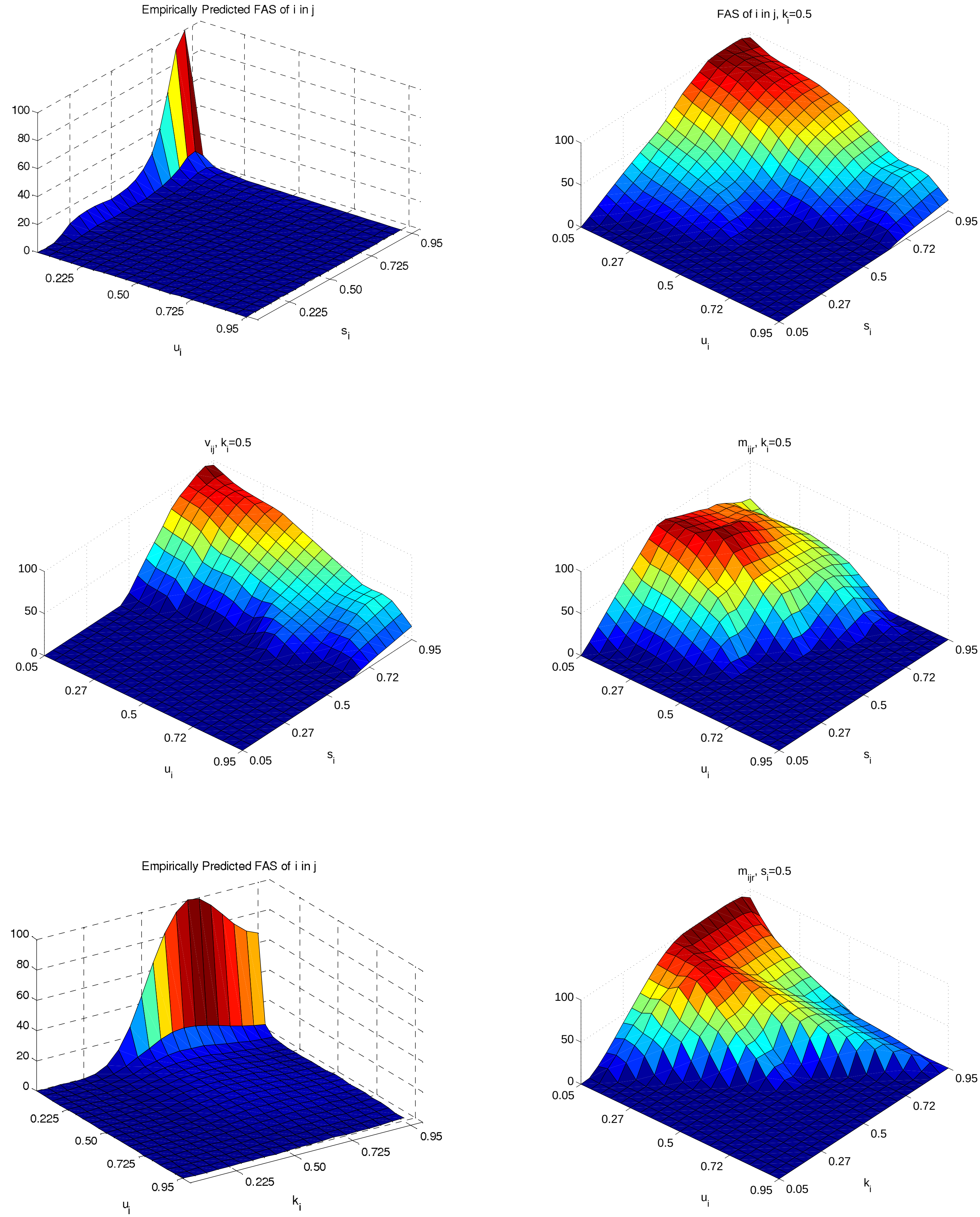




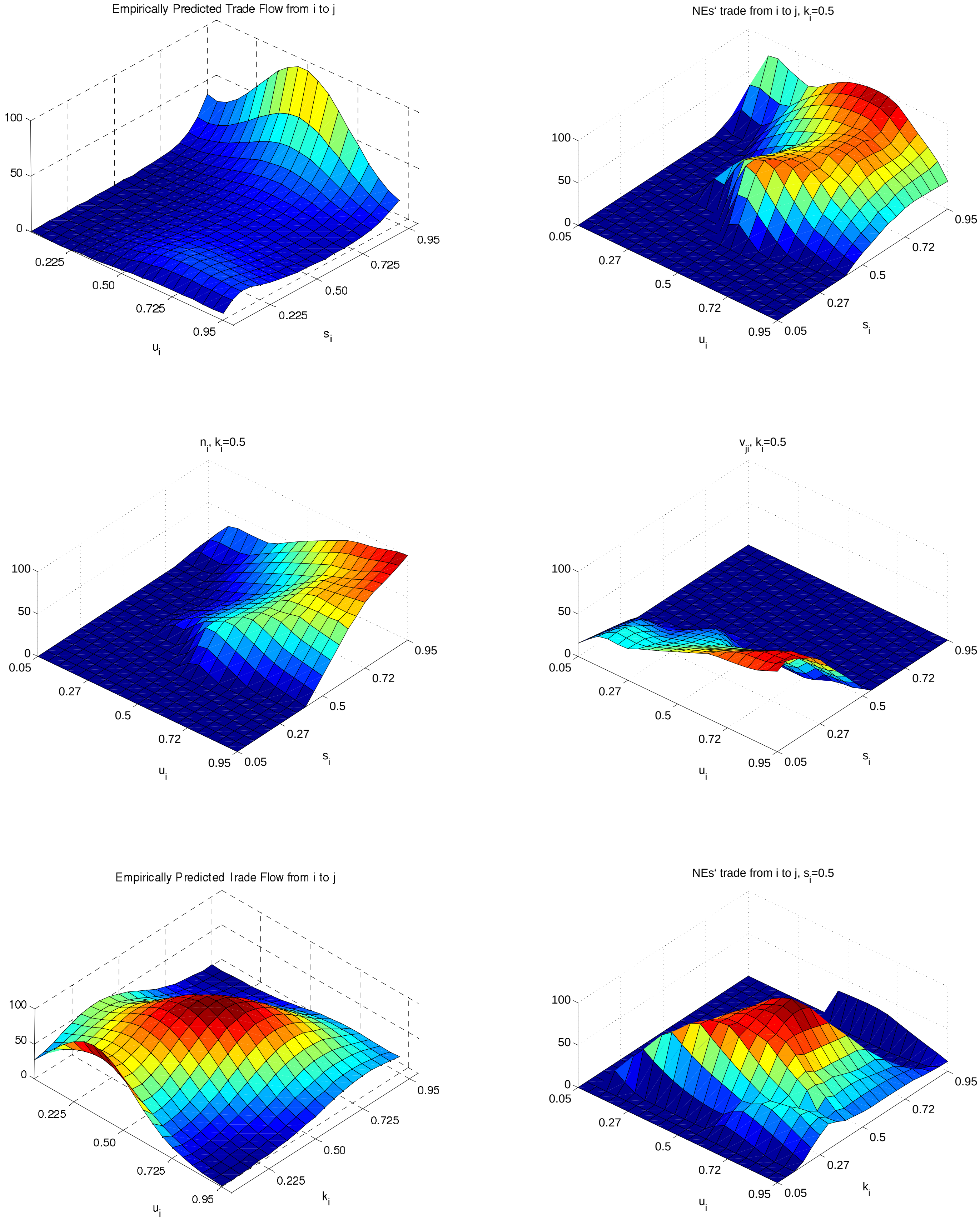




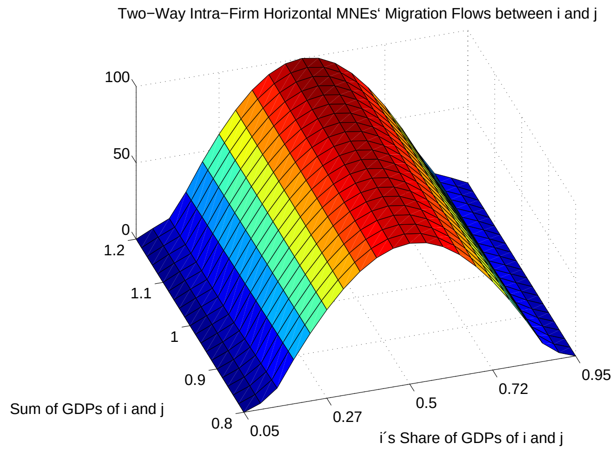
Figures 6a-6f

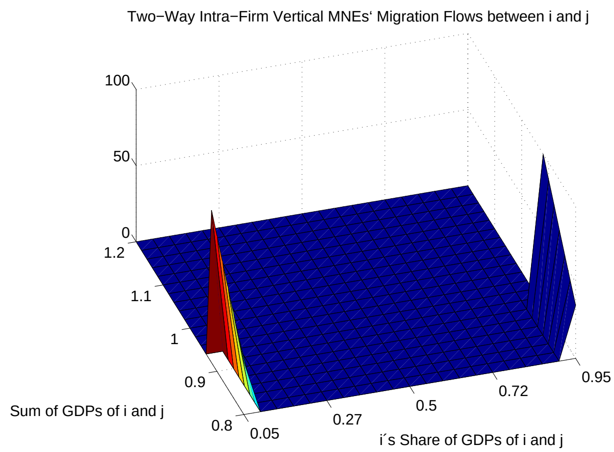


Figures 7a-7f



Figures 8a-8c





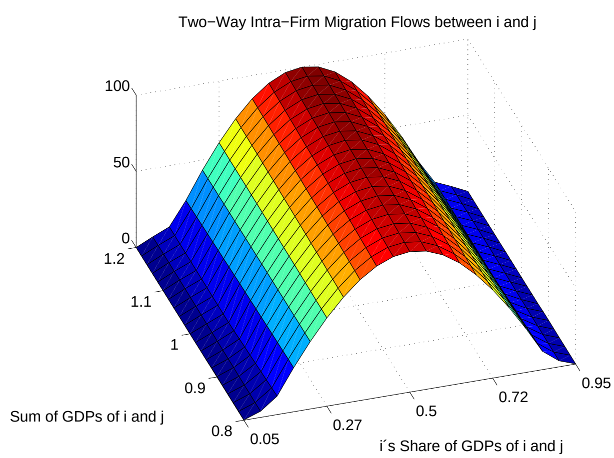


Table 1
Gravity Models for Bilateral Goods Exports, Foreign Affiliate Sales, and Skilled Migrants (Expatriates) from i to j

	Dependent variable is the level of:					
Explanatory variables	Goods Exports$_{ij}$		Foreign Affiliate Sales$_{ij}$		Skilled Migrants (Expat's)$_{ij}$	
log GDP$_i$	0.777	0.157 ^{c}	0.787	0.146 ^{c}	0.591	0.042 ^{c}
log GDP$_j$	0.516	0.120 ^{c}	0.469	0.138 ^{c}	0.880	0.063 ^{c}
Log Distance$_{ij}$	-0.459	0.095 ^{c}	-0.405	0.096 ^{c}	0.017	0.234
Adjacency$_{ij}$	0.329	0.272	0.371	0.265	0.677	0.903
Language$_{ij}$	0.171	0.468	0.222	0.478	1.501	0.326 ^{c}
Constant	3.716	1.686 ^{b}	3.422	1.688 ^{b}	-32.118	2.907 ^{c}
Number of observations	1152		1370		1925	
Pseudo-R^2	0.795		0.414		0.725	
Log pseudo-likelihood × 10^{-6}	-1558.00		-8.45		-3.51	

Notes: Reported coefficients and standard errors are Poisson quasi-maximum likelihood estimates. Standard errors are robust to heteroskedasticity. Superscripts a, b, and c refer to significance levels of 10, 5, and 1 percent, respectively.

Table 2

Poisson Quasi-Maximum Likelihood Estimation of Skilled Migration (of Expatriates) into 95 Economies

Explanatory variables	Coef.	Std. Err.	Explanatory variables continued	Coef.	Std. Err.
K_i+K_j	-	-	$\log(K_i+K_j)$	0.785	0.381 [b]
S_i+S_j	-	-	$\log(S_i+S_j)$	-2.437	0.734 [c]
U_i+U_j	-	-	$\log(U_i+U_j)$	-1.302	0.227 [c]
$GDP_iGDP_j/(GDP_i+GDP_j)^2$	-41.834	17.834 [b]	$\log[GDP_iGDP_j/(GDP_i+GDP_j)^2]$	-	-
k_i	-22.019	29.728	$\log(k_i)$	7.539	19.078
s_i	43.808	58.708	$\log(s_i)$	-71.910	27.108 [c]
u_i	41.306	31.412	$\log(u_i)$	-9.704	39.294
t_i	-0.068	0.812	$\log(t_i)$	-28.021	18.692
γ_i	-0.427	0.662	$\log(\gamma_i)$	35.079	22.147
k_j	-	-	$\log(k_j)$	2.494	0.703 [c]
s_j	-	-	$\log(s_j)$	-2.506	0.658 [c]
u_j	-	-	$\log(u_j)$	-0.729	0.502
t_j	-38.305	5.510 [c]	$\log(t_j)$	1159.315	159.606 [c]
γ_j	16.280	2.771 [c]	$\log(\gamma_j)$	-562.339	100.673 [c]
$k_i\times\log(GDP_i+GDP_j)$	0.607	1.021	$\log(k_i)\times\log(GDP_i+GDP_j)$	-0.418	0.232 [a]
$k_i\times GDP_iGDP_j/(GDP_i+GDP_j)^2$	23.032	14.774	$\log(k_i)\times\log[GDP_iGDP_j/(GDP_i+GDP_j)^2]$	-14.599	7.313 [b]
$k_i\times k_j$	30.655	16.878 [a]	$\log(k_i)\times\log(k_j)$	-19.694	10.751 [a]
$k_i\times s_i$	-	-	$\log(k_i)\times\log(s_i)$	0.055	0.073
$k_i\times s_j$	5.634	7.890	$\log(k_i)\times\log(s_j)$	-0.032	0.794
$k_i\times u_i$	-	-	$\log(k_i)\times\log(u_i)$	-0.013	0.110
$k_i\times u_j$	-6.612	5.129	$\log(k_i)\times\log(u_j)$	-0.341	0.388
$k_i\times t_i$	-0.362	0.267	$\log(k_i)\times\log(t_i)$	3.609	2.020 [a]
$k_i\times\gamma_i$	0.538	0.304 [a]	$\log(k_i)\times\log(\gamma_i)$	-3.554	2.131 [a]
$k_i\times t_j$	0.492	0.674	$\log(k_i)\times\log(t_j)$	1.029	3.962
$k_i\times\gamma_j$	-0.445	0.255 [a]	$\log(k_i)\times\log(\gamma_j)$	0.699	2.418
$s_i\times\log(GDP_i+GDP_j)$	2.975	1.564 [a]	$\log(s_i)\times\log(GDP_i+GDP_j)$	-0.523	0.446
$s_i\times GDP_iGDP_j/(GDP_i+GDP_j)^2$	18.125	17.226	$\log(s_i)\times\log[GDP_iGDP_j/(GDP_i+GDP_j)^2]$	-3.048	5.988
$s_i\times s_j$	-3.714	14.149	$\log(s_i)\times\log(s_j)$	1.168	9.352
$s_i\times u_i$	1.873	6.898	$\log(s_i)\times\log(u_i)$	-0.006	0.117
$s_i\times u_j$	-	-	$\log(s_i)\times\log(u_j)$	-0.495	0.842
$s_i\times t_i$	0.039	0.330	$\log(s_i)\times\log(t_i)$	-4.516	3.595
$s_i\times\gamma_i$	0.104	0.371	$\log(s_i)\times\log(\gamma_i)$	1.637	2.795
$s_i\times t_j$	-36.912	20.141 [a]	$\log(s_i)\times\log(t_j)$	19.037	6.666 [c]
$s_i\times\gamma_j$	-0.206	0.339	$\log(s_i)\times\log(\gamma_j)$	8.923	3.459 [c]
$u_i\times\log(GDP_i+GDP_j)$	-0.718	1.187	$\log(u_i)\times\log(GDP_i+GDP_j)$	0.486	0.340
$u_i\times GDP_iGDP_j/(GDP_i+GDP_j)^2$	-0.201	14.349	$\log(u_i)\times\log[GDP_iGDP_j/(GDP_i+GDP_j)^2]$	1.374	5.042
$u_i\times u_j$	-6.211	13.293	$\log(u_i)\times\log(u_j)$	0.765	9.547
$u_i\times t_i$	1.180	0.523 [b]	$\log(u_i)\times\log(t_i)$	-5.729	4.380
$u_i\times\gamma_i$	-1.023	0.431 [b]	$\log(u_i)\times\log(\gamma_i)$	7.142	3.113 [b]
$u_i\times t_j$	0.502	0.624	$\log(u_i)\times\log(t_j)$	-6.138	8.387
$u_i\times\gamma_j$	-0.787	0.358 [b]	$\log(u_i)\times\log(\gamma_j)$	3.171	5.030
			Other control variables		
$\log(distance_{ij})$	-0.298	0.097 [c]	colonial relatinship$_{ij}$	0.325	0.253
adjacency$_{ij}$	0.666	0.262 [b]	colonial relationship since 1945$_{ij}$	0.442	0.387
common official language$_{ij}$	0.764	0.266 [c]	same country$_{ij}$	0.666	0.510
common spoken language$_{ij}$	0.513	0.185 [c]	constant	-1322.385	185.804 [c]

Notes: The sample consists of 1925 observations. The pseudo-R^2 amounts to 0.910, and the value of the log-likeliood at the reported parameter estimates is -1141201.3.