research paper series

Globalisation, Productivity and Technology

Research Paper 2003/10

Exports and Distance in a Digitized World

by

P.K.M. Tharakan and I. Van Beveren

Leverhulme Centre
for Research on Globalisation and Economic Policy

The Centre acknowledges financial support from The Leverhulme Trust
under Programme Grant F114/BF

The Authors

P.K.M. Tharakan is Professor Emeritus at the University of Antwerp; Ilke Van Beveren is a Research Assistant at the same University.

Acknowledgements

We are indebted to a large number of people for their help and encouragement in carrying out this study. V. Ginsburgh made detailed comments on an earlier version of the paper. T. Van Ourti provided very valuable help by suggesting additional estimation procedures. J. Plasmans, A. Panagariya, J. Tharakan and J. Waelbroeck provided conceptual clarifications. Indian I.T. professionals and managers including Mohan Das Pai (Infosys Ltd.), B.V. Naidu (Director Software Technology Park, Bangalore), and T.A. Varghese (Tata Consultancy Services) spent much time, and patiently answered the numerous questions we posed about the I.T. industry in general and the Indian software sector in particular. K.J. Joseph (Jawaharlal Nehru University), Thomas Abraham (GOPI International), C. Veeramani (Indian Council for Research on International Economic Relations), O. Laveda (International Monetary Fund, IMF), and various people at the National Association of Software and Service Companies (NASSCOM) were of help either in aiding us to obtain the necessary data or in clarifying questions related to it. While we are thankful to all of them, none of them are responsible for any errors we might have made.

Exports and Distance in a Digitalized World: Gravity Model Applied to the Indian Exports of Software

by

P.K.M. Tharakan and I. Van Beveren

Abstract

In this study we apply the gravity model to the fast growing Indian exports of software and to the much more slowly growing total Indian commodity exports. In analyzing the results obtained, we pay special attention to the distance variable, linguistic connections, and trade facilitating networks. The insights gained from our analysis might be relevant in appraising the prospects of the new wave of emerging country exporters of software. Further, the lack of importance of distance as a trade resistance variable in software (which we confirm in our study) has interesting implications for trade liberalization in, and spatial agglomeration of, some industries.

JEL classification: F 14

Keywords: Software, Gravity Model, Distance

Outline

Non-Technical Summary

An overwhelming share of global software production and sale are concentrated in high income countries. But certain emerging economies, especially India, with large pools of skilled and relatively low cost labour have been rapidly increasing their software production, and in some cases, their exports. In this study we analyze the Indian case and draw some specific, and general conclusions from our findings.

We first document the remarkable performance of the Indian software exports. The modes of delivery of Indian exports of software are of three kinds: 1) onsite services, 2) offshore services and 3) products and packages. In the first category, the project is executed at the client's site by Indian software professionals according to the former's requirements and specifications. The offshore services are based on outsourcing by foreign firms. The overwhelming dominance of the combined contribution of the onsite and offshore delivery modes means that India's net export earnings are considerably less – perhaps less than half – of what the gross export figures indicate. The fact that the bulk of Indian software exports consists of rather routine programming and maintenance services has triggered a policy oriented debate in India about the alleged weak linkages of the Indian software sector with the domestic manufacturing industry. Although the global economic downturn has slowed down the growth, the Indian software exports have continued to perform very well due to the increasing need of the outsourcing companies to efficiently use and maximize benefits from their existing infrastructure in the changed economic environment.

The reasons attributed to the remarkable success of the Indian software exports include: the growing demand for software in the countries of the clients; the cost difference between India and the outsourcing countries in employing software professionals; knowledge of English language by the large pool of Indian software professionals, and the possibility to interact with clients from a long distance at a low cost. The role played by people of Indian origin in various countries in facilitating software trade is also sometimes mentioned.

In the econometric part of the study we apply the gravity model to the fast growing Indian exports of software and to the much more slowly growing total Indian commodity exports. In analyzing the results obtained, we pay special attention to: the distance variable, linguistic connections and the trade facilitating networks proxied by the percentage of people of Indian origin in the total population of the importing countries. A number of interesting insights emerge from this exercise. First of all, unlike in the case of total commodity exports, distance does not act as a barrier to the Indian exports of software, which within a few years might account for nearly 20 per cent of India's export earnings. It is reasonable to assume that this lack of significance of distance as an export resistance variable might be equally true for much of the worldwide trade in software, which is sharply increasing its share of world GDP.

Our results also show that the knowledge of the English language by the Indian software professionals has played a significant and positive role in the export performance of this sector. But our results do not permit us to make such an inference in the case of the total Indian commodity exports. The importance of the people of Indian origin in the importing countries show a positive effect in the case of both the Indian software exports and commodity exports, but is consistently significant only for the latter. In general, our results suggest that it will be those countries which are able to make an effective use of a combination of the above-mentioned type of advantages which will emerge as the most successful suppliers of software services to high income markets, irrespective of how far away such markets happen to be.

Our results also have broader implications. For example, if distance does not act as a trade resistance variable in software trade, there is no special reason for favouring regional trade liberalization – as distinct from multilateral trade liberalization – in that sector on the basis of the 'natural partners argument'. On the basis of some variants of the theoretical models dealing with the effect of distance on international trade and international location of economic activity in vertically linked industrial activities, our results indicate that it cannot be automatically assumed that India and the other emerging market exporters now successfully supplying mostly the low-value added segments of software services may easily attract high-value adding downstream investments in the industry.

1. Introduction

The capacity of gravity models to empirically explain bilateral trade between countries has been widely recognized since the publication of the studies in this field of Tinbergen (1962), Pöyhönen (1963) and others. Subsequently, a number of economists proposed theoretical underpinnings which this powerful empirical tool clearly lacked (See Anderson (1979), Bergstrand (1989, 1990), Deardorff (1998), Helpman (1987) for some of the contributions). The recent increase in the recourse to gravity model could be partly ascribed to the growing interest in the new economic geography. The rapidly growing field of new economic geography deals with a much wider set of issues than the role of distance in international trade (see Fuijita and Thisse (2002) for a comprehensive presentation, and Neary (2001) for a critical appraisal, of the field). But it is clear that gravity models which explicitly take into account the role played by distance between the trading countries is an attractive analytical device for empirical analysis of some issues linking trade and space.

In this study we apply the gravity model to the Indian exports of software, and to total Indian commodity exports. As we explain below, the results obtained from the analysis of this particular flow of trade might have more general relevance. In analyzing the results obtained in our exercise, we pay special attention to the distance variable and the variables representing linguistic connections and potential for creating trade facilitating networks. The rationale of the study is explained in this section. A highly condensed case study of the Indian exports of software is given in section 2. The empirical model is presented in section 3, and the results are presented in section 4. In the final section we indicate some of the broader implications of our results.

The software sector, which experienced 100 per cent growth between 1995 and 2001 has been the fastest growing branch in the Information and Communication Technology (ICT) industry, which in turn, according to one estimate, accounted for US$ 2.4 trillion spending, representing 7.6 per cent of the global Gross Domestic Product (GDP) in 2001 (World Information Technology and Services Alliance (WITSA), 2002). While there can be a presumption in favour of the hypothesis that distance does not matter in international trade of products (services) such as software which can be transmitted across the world almost instantaneously, there is no unanimity among specialists on this point. Some scholars (see

Frankel (1997), pp. 46-47) argue that the agglomeration effects seen even in sectors where physical transport costs are negligible (for example, the concentration of software firms in Silicon Valley) show that distance still matters because it can be an impediment to face-to-face contact for exchanging information and negotiating deals. Further, as we will see in the case study (section 2), the predominant mode of export delivery (on-site and/or offshore services) used by emerging market exporters such as India involves travel costs, living allowances abroad, etc. Hence the trade resistance role of distance in software trade needs to be empirically verified, not just assumed away. It is also clear that, in order to meaningfully evaluate the results yielded by the hypothesized determinants of software exports, one has to compare them with the results of a more conventional exercise, which in this case are those obtained by the application of the gravity model to India's total commodity exports.

As Beckerman (1956, p. 38) and Linnemann (1966, pp. 25-34) pointed out long ago, distance has a broader meaning in the context of international trade. In addition to transport costs and transport time, a group of factors, sometimes referred to as 'psychic distance' which facilitate (restrict) bilateral trade might be important. Familiarity with the principal language of the partner country, cultural affinity, etc. could be examples of such trade stimulating elements.

An overwhelming share of the global software production and sale are concentrated in the high income countries with more than 50 per cent of all software sold accounted for by North America (WITSA (2002)). But certain emerging markets such as India, China, Brazil, Mexico and the Philippines with large pools of skilled and relatively low cost labour have been rapidly increasing their software production, and in some cases, their exports. Another group of countries including Hungary, Poland and Russia also have the capability to emerge as major software exporters (see Cane (2002)). Some of these countries appear to have the potential to tap into the trade facilitating networks and linguistic connections, to promote their international trade, particularly that of software. The results of the analysis of the Indian experience could yield interesting insights about the prospects of an important new wave of software exporters entering into this growing flow of international trade.

This study could also help us to make inferences about issues which have broader implications. For example, there is a lively debate between those who hold the view that if distance matters as a trade resistance variable, neighbouring countries would make natural trade partners (Krugman (1991), Summers (1991)), and those who oppose this position (Bhagwati (1992, 1993), Panagariya (1996)). A corollary of the former position is that if distance does not hinder trade in a particular product or service, multilateral trade liberalisation in that sector, as distinct from preferential trade arrangements (PTAs) between neighbours, could well be the preferable order of trade arrangements in economic terms. Again, a recurring theme in some variants of the new economic geography models is that distance, under certain conditions leads to agglomeration. For example, in Venables (1996), the incentive for spatial agglomeration of vertically linked industries is maximal for intermediate levels of the transport costs of inputs and negligible for extreme values, thus following a "U-pattern". If transport costs of inputs are very high, firms will be dispersed in space in order to serve consumers, and if such costs are very low, there will be no incentive for agglomeration. Although observable phenomena can rarely satisfy all the assumptions on which such theoretical models are built, empirical verification of the trade resistance role of distance in specific sectors might help us to make inferences about the possible location pattern of certain industries.

2. Indian Exports of Software[1]

The remarkable growth of the Indian software industry is well documented and analyzed (Arora, Arunachalam, Asundi and Fernandes (2001), Banerjee and Duflo (2000), Heeks (1996), Joseph and Harilal (2001), McKinsey (2001), Nasscom (2002b) are among some of the studies and sources). The Indian software industry is highly export oriented, with the exports amounting to \$ 6.217 billion[2] in Financial Year (FY) 2000-2001 and accounting for about 51.04 per cent of the total revenue of the Indian I.T. sector. The export growth rates averaged above fifty per cent per year during the last decade of the 20th century. Though these rapid growth rates slowed down by FY 2001-2002 (mainly due to the economic

[1] This highly condensed description of the Indian software export sector is based partly on detailed discussions with a number of I.T.-professionals and managers in India (mentioned in the footnote on the title page of this paper), and partly on the case studies cited here.

[2] This includes export revenues from software and services, I.T. enabled services and Research and Development (R&D) services. The expression 'software exports' is used in this paper to cover the above items.

downturn and uncertainty in the main markets), they were still impressive at 23.53 per cent in dollar terms. The exports are service-, rather than product-oriented.

For the present analysis it is useful to distinguish between three categories of Indian software exports on the basis of the modes of delivery used. They are: 1) onsite services, 2) offshore services and 3) products and packages. In the first category, the project is executed at the client's site by Indian software professionals according to the client's requirements and specifications. This involves Indian software personnel moving to foreign locations to carry out the work there. The second category of exports i.e. the offshore services is mainly carried out in India, but it usually involves initial visits to the clients' establishment abroad for the analysis of their requirements, understanding the problem specification, and in some cases for a short period of training. The offshore services involve using high-speed datacom links which allow the use of computers located anywhere in the world by programmers in India on a real-time and on-line basis for servicing the requirements of the clients. The clients, wherever they are located, can monitor the software development on a continuous basis, make quality checks and communicate with the programmers in order to get the required software developed. Organisationally this process could take the form of offshore development centres which work on the basis of long term contracts between the exporters and clients. The work is executed offshore with the exporting (Indian) firm responsible for technical and organisational aspects of the work. More than 200 of the Fortune 1000 leading firms are outsourcing to Indian software firms. The third mode of exports consists of software products and packages.

The Indian software exports are showing a shift towards the offshore mode. Their share in the total Indian software export earnings grew from 34.69 per cent to 38.62 per cent between 1999-2000 and 2000-2001. According to Nasscom estimates, the contribution of offshore earnings exceeded those of onsite services in 2001-2002 and accounted for 49.32 per cent of the total software export earnings in that year[3]. The highly capital- (and marketing-) intensive segment of products and packages has a share of only about 5.30 per cent in 2000-2001. This has declined to 4.10 per cent in 2001-2002. By 2000-2001, the gross exports earnings of the Indian software industry (onsite, offshore and products and

---

[3] Nasscom (2002b)

packages together) accounted for 14 per cent of India's total exports. For the year 2001-2002, the gross export earnings were US$ 7,680 million. A somewhat higher figure (US$ 7,780 million) of exports initially estimated by Nasscom (2002a, p. 28), represented, according to them, 16.3 per cent of the total Indian exports. The overwhelming dominance of the combined contribution of the onsite and offshore delivery means that India's net export earnings from software are considerably less than what the reported figures indicate. Joseph and Harilal (2001, p. 3267) estimate the net earnings to be not more than 50 per cent of the gross export figures, while an earlier estimate by Heeks (1998, p. 2) put that figure at 40 per cent.

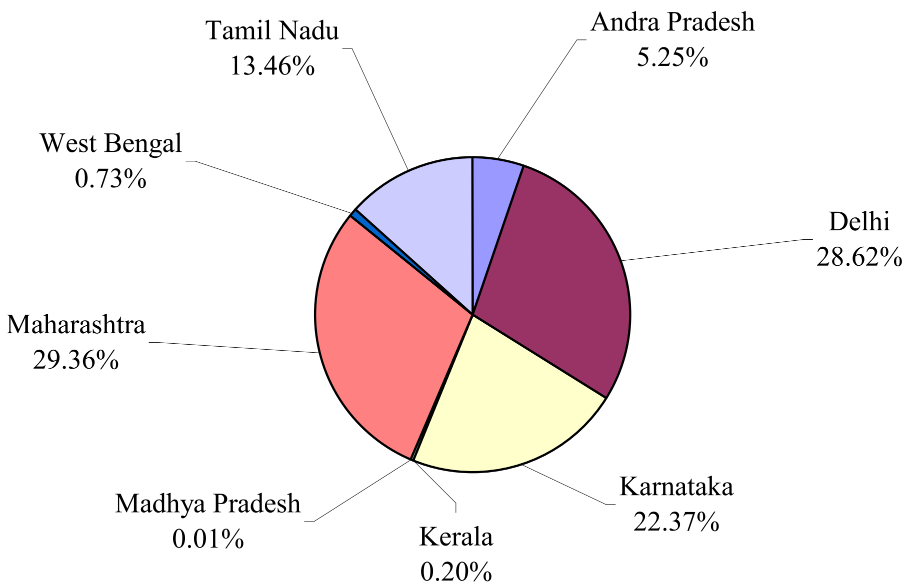
The fact that the bulk of Indian software exports consists of rather routine programming and maintenance services and the failure of products and package exports to move ahead is a cause for some concern (see Heeks (1998), Nath and Hazra (2002)). But the I.T. professionals we interviewed in India are optimistic about the prospects of the sector, particularly in view of the shift towards offshore and higher value added work. Although the bursting of the dotcom bubble and the global economic slowdown have somewhat tempered their enthusiasm, it certainly has not snuffed it out. They point out that the changed conditions would increase the need of companies to efficiently use and maximize benefits from their existing infrastructure. Consequently, integration of applications and their support and maintenance, for which Indian firms are well positioned, will assume greater importance. Further, they also argue that, given the comparative resource endowment patterns and international market conditions, a major push to increase exports of software products and packages will not make economic sense. Nevertheless it is important to note that there is a vigorous debate going on in India at present concerning the alleged weak linkages of the Indian software sector with the domestic manufacturing industry and the possible lack of innovative impact of the structure of Indian exports (see Joseph (2002), Nath and Hazra (2002)). But recent developments such as the decision of Microsoft to invest $ 400 million for software development, software testing, product development and reconfiguring products and applications for specific markets[4] are noteworthy.

[4] K. Merchant (2002).

The Indian software industry is not concentrated exclusively in Bangalore. According to some estimates, already by 1997 (the first year for which data is included in this study) Mumbai and Pune together had a higher number of Nasscom member firms, and a much larger share in the total revenue of Indian software industry than Bangalore (see Arora, Arunachalam, Asundi and Fernandes (2001, p. 1272)). Delhi too had a larger number of Nasscom member firms than Bangalore. Chennai, Hyderabad and Calcutta, in addition to cities in states including Gujarat and Kerala are seeing the development of important software centres. A glance at the map of India will confirm the enormous distance which separates some of these cities.

Chart 1 shows the state-wise share of the origin of the exports realized by the main Indian software exporters during the last year included in our sample (2000-2001). Maharashtra leads with 29.36 per cent of the total software exports. Delhi is a close second with 28.62 per cent. Karnataka (the state where Bangalore is located) comes third with 22.37 per cent. So the idea of a 'silicon plateau' in India is not confirmed by empirical evidence.

Chart 1: State-wise Distribution of Indian Software Exports (2000-01) (% of total)



Source: Indian I.T. Software and Services Directory 2002 (Nasscom (2002c)), based on calculations provided by K.J. Joseph.

The reasons for the remarkable success of Indian software exports have also been analyzed in some of the studies mentioned earlier. The reasons cited by researchers and software professionals include the rapidly growing demand for software in the countries of the clients, the cost difference between India and the outsourcing countries in employing software professionals, knowledge of English language by the large pool of Indian software professionals and the possibility to interact with the clients from a long distance and at a low cost through high-speed datacom links, particularly in the case of offshore development services. The question of the positive role played by the people of Indian origin in various countries in facilitating software trade as well as trade in general with India was also sometimes mentioned. Finally, it should be also noted that trade barriers of the conventional type (tariffs, quotas, etc.) do not affect the trade in software, in contrast to the international trade in commodities. But visa and work permit restrictions faced by the Indian software professionals in some important markets pose difficulties, particularly for the development of onsite services.

3. Empirical Model

3.1. Gravity Model Specification

The gravity model usually expresses a log-linear relationship of bilateral gross trade as a function of certain economic forces in the countries (regions) concerned, and variables resisting and facilitating such trade. Typically, the basic gravity equation can be written as:

$$PX_{ij} = \beta_0(Y_i)^{\beta 1}(Y_j)^{\beta 2}(D_{ij})^{\beta 3}(A_{ij})^{\beta 4}u_{ij} \qquad (1)$$

where PX_{ij} is the value of the trade flow between country i and country j, Y_i (Y_j) represents the nominal Gross Domestic Product of country i (GDP_j), D_{ij} indicates the distance between the economic centers of i and j, A_{ij} refers to any other factor that either aids or resists trade[5], and u_{ij} is the disturbance term.

It is useful to note the following points concerning this simple, basic equation. First, what is sought to be explained here is the gross flow of trade. In other words, this is not an

[5] To simplify presentation, we shall often refer to this set of variables as ‘resistance variables’.

equation which seeks to explain the 'revealed' comparative advantage of the countries concerned. Obviously, this remark is valid whether PX_{ij} is introduced as the value of the total trade between i and j, or whether it is disaggregated to product level.

Secondly, the model assumes that the economic forces at the origin and destination of the trade flow (GDP of country i and GDP of country j) are among the important variables which determine the value (volume) of the flow concerned. Some economists (eg. Tinbergen, 1962, p. 263) interpreted the above variables as representing the economic size of the trading partners. The size of the population of the trading countries (i, j) was not explicitly included in the core model, but later it found its way into the gravity equations either as a separate independent variable or as the denominator of the per capita GDP variable. This last-mentioned variant, when introduced in the form of *differences* between i and j *could be* a proxy for cost differences between the countries concerned (see the following subsection). Thirdly, as was mentioned earlier, an important merit of the gravity model is that it explicitly recognizes the role of distance (D_{ij}) in international trade. Fourthly, the inclusion of the trade resistance variables (A_{ij}) opened up promising areas of investigation, particularly when new theoretical developments gathered momentum years later. Initially this set of variables included, for example, adjacency (in addition to distance)[6], membership in preferential trading arrangements (PTAs), prevalence of the use of the same language, etc. More recently, the role played by co-ethnic networks[7] in influencing bilateral trade, particularly in differentiated products has been stressed (Rauch and Trindade, 2002).

3.2. Specification used

On the basis of the rationale underlying the gravity model and some of its extensions mentioned in the above subsection and the salient facts emerging from the case study presented in section 2, the following regression specification was used to explain India's software exports:

$$X_{ij} = f(GDP_j, GDP_j / POP_j, D_{ij}, LANG_j, PIO_j) \qquad (2)$$

[6] The reasoning here being that adjacent countries might have more intense trade contacts than what can be explained by short distance alone.
[7] We shall often refer to this variable simply as 'networks'.

Where X_{ij}	equals the dollar value of the Indian exports of software to country j
GDP_j	equals the Gross Domestic Product of country j
GDP_j/POP_j	equals the per Capita GDP of country j
D_{ij}	is the Distance between India and country j
$LANG_j$	are countries (j's) where English is widely used (see below for further clarification)
PIO_j	equals the share of People of Indian Origin in the total population of country j (see below for further clarification).

We have then proceeded to test an identical specification for explaining India's total exports (XT_{ij}) for the same years, i.e.:

$$XT_{ij} = f\left(GDP_j, GDP_j / POP_j, D_{ij}, LANG_j, PIO_j\right) \qquad (3)$$

where all the explanatory variables have the same meaning as in equation (2).

Before coming to the expectations concerning the outcome of the explanatory variables in equation (2) and (3), several clarifications are in order. Since one country in each pair of countries (i and j) in our sample is India in all cases, it makes no difference for our analytical purpose whether we multiply the GDP and GDP per capita income variables, or whether we simply use the value of those two variables pertaining to the j's only (the former course of action would amount to multiplying the value of the above-mentioned two variables with a constant). Another interesting implication of the use of GDP and GDP per capita of country j alone, is that they could be interpreted as representing the difference between these two variables pertaining to j and i ($GDP_j - GDP_i$, $(GDP_j/POP_j) - (GDP_i/POP_i)$). In other words , if the differences were to be calculated, we will be deducting a constant figure for all observations for this variable.

As GDP_j represents the economic size of the importing country (or the difference in its economic size relative to India), both in the case of the software exports as in the case of total exports of India, this coefficient is expected to have a positive effect. A similar effect could be expected for GDP_j/POP_j in the case of Indian software exports, because the higher the GDP per capita, the higher could be the demand for software. To the extent that this variable can be seen as a proxy for per capita GDP differences, which are in turn usually correlated with the production costs of software between i and j (as explained in the next paragraph we have some evidence that this could be the case), a similar result could be expected on the basis of the cost effectiveness of the Indian software professionals. Given the fact that the Indian software sector operated with much less policy-induced distortions than India's industrial sector in general, such an interpretation is clearly admissible for the software sector (see section 4 for a clarification of this point)[8].

Detailed comparison between the remuneration earned by the software professionals in India and her trading partners is not possible due to the lack of availability of data covering all the countries involved. Comparable data pertaining to the wages of software professionals are available to us only for the following countries: Canada, Greece, India, Ireland, Switzerland, United Kingdom and the United States. They are for the year 1995 (Heeks, 1999). This information pertains to the following categories of software professionals: project leader (a), systems analyst (b), business analyst (c), systems designer (d), development programmer (e), support programmer (f), network analyst / designer (g), quality assurance specialist (h) and database analyst (i). Given the limited number of degrees of freedom available, we have resorted to non-parametric tests to verify the rank correlation between GDP per capita and the wages to each one of the above-mentioned nine categories of software personnel. In every case the Spearman rank correlation coefficient was positive and significant at the one per cent confidence level.

The distance variable (D_{ij}) consists of the great circle distance between the latitude – longitude combinations of the capitals of countries i and j (for the specifics of the calculation of this variable, see appendix 1). The critiques against this way of quantifying the distance variable, and the responses to those critiques are well known (see Linnemann (1966, pp. 25-31), Frankel (1997, pp. 65-70) for a summing up of the main points concerning this question). While they do not need any reiteration here, an interesting and related conceptual point raised by some economists (see for example Polak (1996), Deardorff (1998), Panagariya (1998)) does. Exports by country i to country j depend not just on i's distance from j (D_{ij}), but also on a measure of i's average distance from all other

[8] The prediction of certain key parameter values, particularly those related to GDP is made complex also by the type of (alternative) theories which might underlie the empirical phenomenon examined (see Feenstra, Markusen and Rose (2001) for an illustration of this point).

trading partners. Hence, bilateral distance in gravity-type equations needs to be corrected for average distance. But statistically this argument will only hold if the sample contains more than one i and j. Our sample has one i and many j's, hence weighting the bilateral distance from i's trading partners with the average distances between all pairs in the sample statistically makes no difference.

If in the digitalized world of software transactions, physical distance is not an important trade resistance factor any more, the D_{ij} variable *will not yield a significant result* in the regressions for India's software exports. But, within the logic of the gravity model, the same variable should yield a negative and significant coefficient for the regressions for India's total exports.

English is the principal language in which software services are rendered in the main markets in the world. India is believed to have the second largest group of English-speaking scientific professionals (second only to that of the U.S.) (Nasscom, 2001a, p. 43). Given the familiarity of India's software professionals with the English language, software export transactions with English-speaking countries are likely to be facilitated[9]. Hence we expect a positive sign for the language (English) dummy[10] in the case of India's software exports, but we are less sure about such an outcome for India's total commodity exports as the knowledge of English language does not bestow the same degree of advantage to the more geographically diversified commodity exports.

Gould (1994), Rauch and Casella (1998), and Rauch and Trindade (2002) have emphasized the trade facilitating role played by networks. One of the channels through which such groups might promote bilateral trade is by providing market information and by supplying matching and referral services. People of Indian origin (PIOs) living abroad probably belong to more heterogeneous groups than immigrants from other countries and hence it is

---

[9] But given the intention of the Indian software industry to geographically diversify its exports, and in view of the increased efforts by the Indian I.T. professionals to acquire working knowledge of other foreign languages, the importance of this variable (English) could decline in the future.

[10] Hutchinson (2002) carries out an interesting exercise in which the number of people who speak English as a first language or English as a second language is used, which makes a logarithmic transformation of that variable possible. The need to match our data set without losing too many observations made us opt for the choice of the dummy variable.

not certain if the above-mentioned mechanism operates in the Indian case. In the case of certain commodities, the demand emanating from a large number of PIOs might be a positive factor. Anecdotal evidence suggests that people of Indian origin might have been active in facilitating bilateral software transactions, particularly those related to the U.S. (see Arora, Arunachalam, Asundi and Fernandes (2001, p. 1271)). The Silicon Valley Indian Professionals Association (SIPA) and the Indus Entrepreneur (TiE) are just two examples of the above-mentioned type of networks in software (see Saxenian, 1999). The role played by the Indian (mainly Gujarati) community in Antwerp in building up India's highly successful exports of cut and polished diamonds to the European Union (E.U.) is a similar example for a different sector.

In the study of Rauch and Trindade (2002), one of the two different variables which was taken into account for a similar purpose was the product of the ethnic Chinese population shares for countries i and j. The rationale for this is that this measure indicates the probability, if we select an individual at random from each country, that both will be ethnic Chinese. The recently published Report of the High Level Committee on the Indian Diaspora (Ministry of External Affairs, Government of India (2002)) provides, for the first time, detailed country-wise estimates of People of Indian Origin. Unfortunately a breakdown of these figures by the educational level of the PIOs is not available. We have used the share of PIOs in the total population of the host countries (j's) in the regressions for India's software exports and total commodity exports, although in that particular form our PIO_j variable is more appropriate for the latter than for the former specification. While we expect a positive sign in both instances, we will interpret the results for the PIO_j variable in the software regressions with caution.

Preferential Trading Arrangements (PTAs) and Adjacency have often figured as explanatory variables in gravity models. Neither is relevant in the present analysis. The South Asian Association of Regional Cooperation (SAARC) of which India is a member has been particularly unsuccessful in becoming operational. India's trade with the main adjacent countries is troubled by political tensions and the value of the trade with them and also with Nepal, Myanmar, Bangladesh and Bhutan is negligible[11].

---

[11] Only two of the countries from SAARC and from the adjacent category even figure in the samples.

Specifically, the regression equations were written as:

$$\log X_{ij} = \alpha_0 + \beta_1 \log GDP_j + \beta_2 \log(GDP_j / POP_j) + \beta_3 \log D_{ij} + \beta_4 LANG_j + \beta_5 \log(PIO_j) + \varepsilon \tag{4}$$

$$\log XT_{ij} = \alpha_0 + \beta_1 \log GDP_j + \beta_2 \log(GDP_j / POP_j) + \beta_3 \log D_{ij} + \beta_4 LANG_j + \beta_5 \log(PIO_j) + \varepsilon \tag{5}$$

We have first carried out pooled Ordinary Least Squares (OLS) regressions[12] (with year-specific intercepts) which include the data pertaining to the years 1997, 1998, 1999 and 2000[13]. Subsequently we ran separate regressions for each one of the above-mentioned years. As Appendix 2 shows, collinearity between the independent variables is low. While we have given the sample correlations for the pooled software data (in order to save space), it is representative for the situation for all the years. For only one individual year (2000), the simple correlation between any two independent variables in the software sample reached 0.44. The multiple R^2 method for detecting multicollinearity when applied to the pooled software sample, with distance as the dependent variable, yielded a value of 0.39. The corresponding figure in the case of the sample used for India's total commodity exports is 0.19. The summary statistics of the data used in the pooled regression (for software exports) are given in Appendix 3. The data sources can be found in Appendix 5.

4. Empirical Results

Given the fact that our analysis covers a number of regressions, it is necessary to streamline the process by indicating the direction and sequence which the analysis will follow. After

---

[12] We considered the possibility of including all the '0 value observations' of the dependent variable in our sample and carrying out a logit or tobit analysis. But given the fact that the development of India's software exports is a recent phenomenon, the absence of exports to some countries cannot be considered as a result of the effect of the independent variables in our empirical model. In the case of total commodity exports all the countries for which observations could be matched were taken into account.

[13] Note that the dependent variable pertains to the financial years ending on the 31st of March (for example, what we consider as the export figures for 1997 in fact refers to those for the financial year starting on the 1st April 1997 and ending on 31st March 1998, and so

some comments about the summary statistics and the quality of regression results in general, we report in table 1 the results obtained for the pooled (for the years 1997 to 2000) regressions. Subsequently we present the results for each one of the above-mentioned years. In our analysis we shall concentrate our attention on the performance of the distance variable, comparing and contrasting the results obtained for the Indian exports of software with those obtained for the total Indian commodity exports. We then analyze the results of the two trade-facilitating variables: language and networks. Finally we take up the analysis of the results of the other variables in the specification. In order to facilitate the comparison between the regression results pertaining to the software exports and those obtained for the total commodity exports, we have reported them side by side in every table. All variables except the Language dummy were introduced in the regressions in logarithmic form.

As we have already explained in section 3.B., for the correlation between the independent variables, the highest value (for the software sample) occurs in the case of GDP – GDP per capita for which it varies between 0.39 (1998) and 0.44 (2000). Neither do we confront the problem of heteroskedasticity, except for the regressions pertaining to the year 1997[14]. The coefficient of determination adjusted for degrees of freedom ($\overline{R^2}$) in all regressions for India's total commodity exports, except the one for the year 1997, is above 0.80. For the Indian exports of software they are lower, but still above 0.63 for all regressions except for the year 2000.

In the results reported in table 1, the distance (capital to capital) variable (D_{ij}) yields a negative sign and is highly significant in the regressions for India's total commodity exports. The coefficient has a value of –0.652. In general terms this result could be interpreted as indicating that, controlling for the other variables, the increase of the distance between India and an importing country by 1 per cent, decreases India's total commodity exports by about 0.65 per cent. So distance clearly plays a significant role as an export resistance variable as far as India's total commodity exports are concerned.

---

forth). This means that in the case of the first two independent variables (GDP_j and GDP_j/POP_j) a lag of three months exists.

[14] White heteroskedasticity-consistent standard errors and variance were used in this case.

But the above-mentioned effect does not operate in the case of India's software exports. In the regressions for software, the log distance variable yields a coefficient which does not approach any acceptable level of significance. So in sharp contrast to India's total commodity exports, the trade resisting effect of distance disappears in the case of India's software exports.

Table 1: OLS Regression Results: Pooled Sample (1997-98-99-2000) (Distance: between New Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX)		log(TOTEX)	
n	213		459	
Adj. R^2	0.638404		0.821412	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant-1997	-6.735614	-4.570565*	0.670940	0.843232
Constant-1998	-6.570920	-4.474852*	0.933264	1.166070
Constant-1999	-6.241829	-4.261592*	0.914897	1.149922
Constant-2000	-6.082575	-4.149424*	1.078631	1.368642
$\log(GDP_j)$	0.493571	8.436764*	0.936023	37.30833*
$\log(GDP_j/POP_j)$	0.306601	3.773741*	-0.301343	-8.003445*
log(Distance)	-0.109827	-0.603326	-0.651817	-6.673913*
Language	1.345086	5.375766*	0.051008	0.427818
log(PIO-share)	0.086503	2.321790*	0.189529	11.26751*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

We have confirmed the robustness of the above-result by introducing the variable D_{ij} as the distance between the principal economic centres of the countries concerned. As can be seen from Appendix Table 4.1, which shows the regression results obtained for the pooled sample (using the alternative distance variable), the pattern remains the same. The coefficient attached to D_{ij} has a value of –0.861 and a t-statistic of –9.243 in the regressions for India's total commodity exports. The corresponding values for D_{ij} in the regression for the Indian exports of software were –0.053 and –0.261 respectively.

The English language dummy ($LANG_j$) yields a positive and highly significant coefficient in the pooled regressions for the Indian exports of software. This result is very much in accordance with the evidence that emerged from the case study. The endowment of a large stock of I.T. professionals proficient in English is an important determinant of India's success in software exports. The picture is not the same in the case of India's total commodity exports. The language variable does yield a positive coefficient but its level of significance is weak (t-value of 0.428). Given the greater geographical dispersion of India's total commodity exports, this is a reasonable outcome.

The share of People of Indian Origin in the total population of the importing countries (PIO_j) yields a highly significant result in the case of the regressions for India's total commodity exports. The value of the coefficient is low (0.190). In the case of the regressions for the Indian exports of software, the result is positive and the coefficient significant, but its value (0.087) is lower than in the case of the regressions for India's total commodity exports. Thus it could well be that People of Indian Origin in the importing countries have played a role in facilitating India's software exports. But as we will see below, the results yielded by this variable showed a lack of robustness in the separate regressions carried out for one of the years and in the random effects estimates.

The remaining two variables of the gravity model are the GDP and GDP per capita of the countries to which India exports. The GDP variable yields a positive and highly significant result in the case of both software exports and total commodity exports. In the case of the latter the coefficient is higher (0.936) than in the case of the former (0.494). The GDP per capita results differ for the two regressions. It yields a positive and significant result in the regression for the software exports. In accordance with what we noted in 3.B, this could be interpreted in two complementary ways. First, it is clearly probable that the high demand for software emanating from the world's rich countries has substantially contributed to the growth of the Indian exports of software. The second plausible interpretation of the result yielded by the (GDP_j/POP_j) variable is that the cost difference between India and the importing countries – to the extent the per capita GDP difference reflects such differences – contributes positively in favour of the Indian exports of software. But such an effect does not appear in the regressions for India's total commodity exports. The GDP per capita variable yields a negative sign and is highly significant. We have alluded a possible reason for this result in 3.B.

Nevertheless, in view of the rather unconventional nature of the results obtained for the GDP per capita variable in the regressions for India's total commodity exports, we have explored the use of another gravity model compatible variant of specifications (2) and (3). As was mentioned in subsection 3.A., instead of per capita GDP, the population of the trading countries i and j (or the difference between the two) is sometimes used as an explanatory variable in some empirical exercises. The rationale here could be that the population of a country represents a different aspect of its size and hence a positive sign could be expected. The results obtained by some authors (see Hutchinson (2002)) are in accordance with this interpretation. It is evident that while the replacement for GDP_j/POP_j by POP_j (population of country j) might be reasonable in the case of the specification for India's total commodity exports, it is better to fine-tune this variable in the case of the specification for software exports. We have done this by using the number of personal computers per thousand people in country j ($PCPOP_j$) in the modified variant of specification (2). Note that the highest simple correlation amounts to 0.367 (between GDP_j and $PCPOP_j$) for the pooled software sample and to 0.301 (between GDP_j and POP_j) for the pooled total commodity exports. The regression results of the modified specifications are reported in table 4.2 in the Appendix. In general, the pattern of the results for the variables used in the former specifications remains the same. For the new variables, the outcome is as follows: In the regressions for the Indian exports of software, the variable representing the number of computers per population yields a positive and significant coefficient. In the case of the regressions for the total commodity exports of India, the population variable also yields a positive and significant result.

Given the fact that our data set consists of a cross-section of countries which are likely to have characteristics specific to them (other than those considered relevant in our model), it was thought useful to carry out random effects Generalized Least Squares (GLS) estimations[15] in order to verify if the results would be similar to those obtained in the pooled OLS regressions reported in table 1. The outcome of this exercise is reported in Appendix Table 4.3. The close similarity of the two sets of results is evident. There is one difference which consists of the fact that in the random effects GLS regressions for the software exports, the PIO variable does not yield a significant coefficient. The lack of robustness of the result for this variable in the software regressions has been already alluded to and the reasons behind it will be elaborated below. Note that distance remains insignificant in the regressions for software, while it has yielded a negative and significant coefficient in the regressions for the total commodity exports, thus confirming the outcome of the OLS regressions. In the remainder of our analysis we will confine ourselves to the OLS regressions.

[15] For the appropriateness of the use of the random effects GLS regressions and the method used, see Verbeek (2000).

In table 2 (A, B, C and D) we have reported the regression results for specifications (2) and (3) for the individual years. Generally they follow a similar pattern with the results pertaining to the year 2000 being something of an exception.

Table 2: Regression Results for Individual years

Table 2A: OLS Regression Results for 1997 (Distance: between New Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX97)		log(TOTEX97)	
n	56		119	
Adj. R^2	0.670046		0.792043	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant	-6.161492	-2.865136*	-0.426531	-0.248670
log(GDP_j)	0.446134	5.248792*	0.962663	18.30454*
log(GDP_j/POP_j)	0.195603	1.690784*	-0.359662	-4.077245*
log(Distance)	-0.001574	-0.005988	-0.519941	-2.388340*
Language	1.142332	3.249544*	-0.057214	-0.219254
log(PIO-share)	0.104780	1.742292*	0.234076	5.923046*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

Table 2B: OLS Regression Results for 1998 (Distance: between New Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX98)		log(TOTEX98)	
n	52		117	
Adj. R^2	0.656961		0.805231	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant	-7.532739	-2.801617*	0.787496	0.591987
log(GDP_j)	0.535919	5.227685*	0.909457	16.38767*
log(GDP_j/POP_j)	0.266891	1.921542*	-0.286352	-3.155445*
log(Distance)	-0.043637	-0.135655	-0.608866	-3.672666*
Language	1.501066	4.124484*	-0.038536	-0.164492
log(PIO-share)	0.117499	1.737325*	0.185010	5.328309*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

Table 2C: OLS Regression Results for 1999 (Distance: between New Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX99)		log(TOTEX99)	
n	51		110	
Adj. R^2	0.711203		0.832586	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant	-8.583761	-3.160509*	0.533151	0.403663
log(GDP_j)	0.565698	5.596061*	0.948804	16.82526*
log(GDP_j/POP_j)	0.282918	2.019711*	-0.283544	-3.178302*
log(Distance)	0.042451	0.129094	-0.632066	-3.782122*
Language	1.316417	3.483700*	0.157325	0.654538
log(PIO-share)	0.156174	2.396598*	0.175806	4.995231*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

Table 2D: OLS Regression Results for 2000 (Distance: between New Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX00)		log(TOTEX00)	
n	54		113	
Adj. R^2	0.491398		0.845455	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant	-6.285819	-1.535211	2.873768	2.323211*
log(GDP_j)	0.466003	3.445342*	0.925617	18.11957*
log(GDP_j/POP_j)	0.451632	2.340065*	-0.276015	-3.362164*
log(Distance)	-0.123341	-0.221435	-0.868941	-5.675776*
Language	1.267361	2.229402*	0.179881	0.849279
log(PIO-share)	0.002456	0.028265	0.159229	5.042300*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

In the regressions for each one of the years, the distance variable yields a negative and significant result for India's total commodity exports. But for none of those years the distance variable shows a significant correlation with India's software exports. So the finding that unlike in the case of commodity exports, distance does not act as a trade resisting variable for India's software exports is a robust one. In the additional regressions (not reported here) which we carried out by using the distance between the principal economic centres of the trading partners as the D_{ij} variable, the same pattern of results was seen. Distance had a negative and significant effect on India's total commodity exports for every year taken into consideration in our study. But the estimated coefficient for the same variable was not significant for any of the years in the regressions for software exports.

The highest t-value obtained for the D_{ij} variable in the software regressions with the alternative estimate of distance was 0.25 (for the year 1999).

The finding that in spite of the large share of the onsite mode of delivery of Indian exports of software (which involves travel costs and living expenses abroad for I.T. professionals) distance plays no trade resisting role is clearly important. As the share of the offshore mode of delivery increases in India's software exports, the lack of importance of distance as a 'natural barrier' will become even more explicit in this most dynamic sector of India's foreign trade.

The difference in the results obtained for the English language dummy in the software regressions and the total commodity exports regressions noted in the pooled regressions remains valid for every one of the years taken into account in our study. The coefficient attached to that variable is positive and highly significant for the software regressions for every year. The value of that dummy variable which reached its highest point in 1998 showed a slight decline during the next two years. In the case of the regressions for India's total commodity exports, the language variable is not significant for any of the years included in our sample.

The performance of the variable representing the share of People of Indian Origin in the importing countries (PIO_j) does not show the type of robustness shown by the distance variable or the English language dummy. As in the case of the total commodity exports, PIO_j yields a positive and significant result in the software regressions for 1997, 1998 and 1999. Inexplicably it loses significance in the regressions pertaining to the software exports of 2000. Careful verification of the data, the different steps in the calculations and the correlation matrices does not give us any definitive answer for this change. It is true that the rate of growth of the Indian exports of software that particular year was somewhat higher than the average for the sample period (57.05 per cent compared to 52.01 per cent). But then the rate of growth of the total commodity exports for the year 2000 was also higher than the average for the sample years (15.76 per cent compared to 10.04 per cent). As was alluded to earlier (see section 3.B), we can only speculate that a narrower (and more appropriate) quantification of the PIO_j variable probably would have given a more robust result.

The GDP variable yields a positive and significant result in the year by year regressions for both the software exports and total commodity exports. There is no surprise here. The GDP per capita variable also yields a positive and significant result for the regressions for India's exports of software in every year in our sample. But as in the case of the pooled regressions, the GDP per capita variable's result is negative and significant in the case of total commodity exports. In analysing the similar result obtained for this variable in the pooled regressions, we have indicated our interpretation of this outcome.

5. Concluding Remarks

A number of interesting points emerge from our application of the gravity model to the Indian exports of software. First of all, unlike in the case of total commodity exports, distance does not act as a barrier to the Indian exports of software. This is particularly significant because within a few years, perhaps close to 20 per cent of the total value of India's export earnings could be contributed by the software sector. Further, it is only reasonable to assume that the lack of significance of distance as an export resistance variable, which we have confirmed in the case of the Indian exports of software, might be equally true for much of the worldwide trade in software which is sharply increasing its share of the world GDP.

The disappearance of distance as a trade resistance factor for a product (service) for which certain low income countries endowed with large stocks of technically skilled workers might have genuine cost advantage, has interesting implications. The most obvious one is the likelihood that in the absence of natural (and artificial) trade constraints, such countries will tend to increasingly specialize in the above-mentioned type of sectors. Their advantage is of an absolute-, rather than of a comparative nature[16]. But in addition to the competitive edge provided by absolute advantage in promoting exports, the modern

[16] For example, in a two country, two factor, two product model, a country like India with a relatively high endowment of unskilled labour is more likely to have a comparative advantage in unskilled labour-intensive products than in skilled labour-intensive software. This statement has to be qualified by the observation that *within the software sector itself* India has so far specialized in the comparatively less-skilled activities.

business environment associated with the software sector and the positive externalities which it could generate has the potential to help sustain the upward trend already noticed.

Although India was one of the first low income countries to realize that it had the endowments which allowed it to offer competitive software development services abroad, a number of other countries are now emulating India. Some of them possess assets such as a large pool of low-cost, highly skilled software professionals, working knowledge of internationally used languages (especially English), and the potential for trade facilitating contacts through the presence of earlier waves of emigrants in industrialized countries. It will be those countries which are able to make effective use of a combination of these advantages that will emerge as the most successful suppliers of software services to the high income markets, irrespective of how far away such markets happen to be.

Another point of interest is the implication for regional trade liberalization schemes which have grown in number in recent years. If the effect of distance is found to be insignificant for exports in a given sector, an important, and much debated claim about the 'natural partners' for regional integration in that sector disappears. Recall that in the case studied here, distance is not highly correlated with other independent variables which include language, cultural ties, etc. While the last-mentioned variables play a positive and significant role in India's exports of software, their presence (occurrence) shows no regional concentration. This strengthens the argument of those who hold the view that if regional trade arrangements can bring about gains to the participants, multilateral trade liberalization can yield even greater gains, at least in certain sectors.

A number of theoretical models have analyzed the effect of distance (or transport costs) and other variables on international trade and international location of economic activity in vertically linked industries. In some variants of such models, the impetus for spatial industrial agglomeration will be negligible if the transportation costs of the inputs are very high, or very low. The Indian software exports studied here could be seen mostly as an intermediate product used in the preparation of the final product or service. Distance is apparently not a constraint for the exports in this particular sector, although it clearly plays a trade resistant role in the much more slowly growing commodity exports of India. So it cannot be automatically assumed that India and other emerging market exporters now successfully supplying the low-value added segments of software services may easily attract high-value adding downstream investments in the industry. Other factors such as the abundance of high-skilled, low-cost labour which could be advantageous in the downstream activities also, rather than distance, could be the important determinants of the development of a vertically integrated industry in this sector.

REFERENCES

Anderson, James E., "A Theoretical Foundation for the Gravity Equation," *American Economic Review* 69:1 (1979), 106-116.

Arora, Ashish, V.S. Arunachalam, Jai Asundi, and Ronald Fernandes, "The Indian Software Services Industry," *Research Policy* 30:8 (2001), 1267-1287.

Banerjee, Abhijit V. and Esther Duflo, "Reputation Effects and the Limits of Contracting: a Study of the Indian Software Industry," *Quarterly Journal of Economics* 115:3 (2000), 989-1017.

Beckerman, W., "Distance and the Pattern of Intra-European Trade," *Review of Economics and Statistics* 38 (1956), 31-40.

Bergstrand, Jeffrey H., "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade," *Review of Economics and Statistics* 71:1 (1989), 143-153.

Bergstrand, Jeffrey H., "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade," *The Economic Journal* 100: 403 (1990), 1216-1229.

Bhagwati, Jagdish, "Regionalism versus Multilateralism," *The World Economy* 15:5 (1992), 535-555.

Bhagwati, Jagdish, "Regionalism and Multilateralism: An Overview" (pp. 22-57), in Jaime de Melo and Arvind Panagariya (Ed.), *New Dimensions in Regional Integration* (New York: Cambridge University Press, 1993).

Cane, A., "Russia's One Big Chance to Play Catch Up," *Financial Times* p. 2 (2002), 4th September.

Central Bank of China (Taiwan), *Financial Statistics* (Taiwan, Central Bank of China, 2002).

Deardorff, Alan V., "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" (pp. 7-28), in Jeffrey A. Frankel (Ed.), *The Regionalization of the World Economy* (Chicago: University of Chicago, 1998).

Feenstra, Robert C., James R. Markusen and Andrew K. Rose, "Using the Gravity Equation to Differentiate among Alternative Theories of Trade," *Canadian Journal of Economics* 34:2 (2001), 430-447.

Frankel, Jeffrey A., *Regional Trading Blocs in the World Economic System* (Washington, D.C.: Institute for International Economics, 1997).

Fuijita, Masahisa and Jacques-François Thisse, *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth'* (Cambridge: Cambridge University Press, 2002).

Getty Research Institute, *The Getty Thesaurus of Geographic Names* (www.getty.edu: Getty Research Institute, 2001).

Gould, David M., "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *Review of Economics and Statistics* 76:2 (1994), 302-316.

Government of India, *Report of the High Level Committee on Indian Diaspora* (www.indiandiaspora.nic.in: High Level Committee on Indian Diaspora, Ministry of External Affairs, Government of India, 2002).

Heeks, Richard, *India's Software Industry: State Policy, Liberalisation and Industrial Development* (London: Sage Publications Ltd., 1996).

Heeks, Richard, "The Uneven Profile of Indian Software Exports," *Development Informatics Working Paper Series* No. 3 (1998).

Heeks, Richard, *Indian Software Labour: Cost Breakdown and Comparison* (www.idpm.man.ac.uk/ idpm/isicost: University of Manchester, 1999).

Helpman, Elhanan, "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," *Journal of the Japanese and International Economics* 1:1 (1987), 62-81.

Hutchinson, K., "Does Ease of Communication Increase Trade? Commonality of Language and Bilateral Trade," *Scottish Journal of Political Economy* 49:5 (2002), 544-556.

IMF, *International Financial Statistics* (CD-rom: International Monetary Fund, 2001).

IMF, *Direction of Trade Statistics Yearbook* (Washington D.C.: International Monetary Fund, 2001).

Joseph, K.J. and K.N. Harilal, "Structure and Growth of India's IT Exports: Implications of an Export-Oriented Growth Strategy," *Economic and Political Weekly* 36:34 (2001), 3263-3270.

Joseph, K.J., "Growth of ICT and ICT for Development: Realities of the Myths of the Indian Experience," *WIDER Discussion Paper Series* No. 2002/78, 2002.

Krugman, Paul, “The Move Toward Free Trade Zones,” *Policy Implications of Trade and Currency Zones* (Wyoming: Proceedings of symposium sponsored by the Federal Reserve Bank of Kansas City, 1991).

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

McKinsey, *India: The Growth Imperative* (www.mckinsey.com: McKinsey Global Institute, 2001).

Merchant, Khozem, “Microsoft to Invest $ 400 Million in Indian I.T. Projects,” *Financial Times* p. 22 (2002), 13th November.

Nasscom, *The I.T. Software and Services Industry in India: Strategic Review 1999* (New Delhi: National Association of Software and Service Companies, 2000).

Nasscom, *The I.T. Software and Services Industry in India: Strategic Review 2000* (New Delhi: National Association of Software and Service Companies, 2001a).

Nasscom, *The I.T. Software and Services Industry in India: Strategic Review 2001* (New Delhi: National Association of Software and Service Companies, 2001b).

Nasscom, *The I.T. Software and Services Industry in India: Strategic Review 2002* (New Delhi: National Association of Software and Service Companies, 2002a).

Nasscom, *Nasscom FY 2002 Software and Service Industry Performance* (www.nasscom.org: National Association of Software and Service Companies, 2002b).

Nasscom, *Indian I.T. Software and Services Directory 2002* (New Delhi: National Association of Software and Service Companies, 2002c).

Nath, P. and A. Hazra., "Configuration of Indian Software Industry," *Economic and Political Weekly* 37:.8 (2002), 737-742.

Neary, Peter J., “Of Hype and Hyperbolas: Introducing the New Economic Geography,” *Journal of Economic Literature* 39:2 (2001), 536-561.

Panagariya, Arvind, “The Free Trade Area of the Americas: Good for Latin America?,” *The World Economy* 19:5 (1996), 485-515.

Panagariya, Arvind, “Do Transport Costs Justify Regional Preferential Trading Arrangements? No,” *Weltwirtschaftliches Archiv* 134:2 (1998), 280-301.

Polak, Jacques J., “Is APEC a Natural Regional Trading Bloc? A Critique of the ‘Gravity Model’ of International Trade,” *World Economy* 19:5 (1996), 533-544.

Pöyhönen, Pentti, “A Tentative Model for the Volume of Trade Between Countries,” *Weltwirtschaftliches Archiv* 90:1 (1963), 93-100.

Rauch, James E. and Allessandra Casella, "Overcoming Informational Barriers to International Resource Allocation: Prices and Group Ties," *National Bureau of Economic Research Working Paper Series* No. 6628 (1998).

Rauch, James E. and Vitor Trindade, "Ethnic Chinese Networks in International Trade," *Review of Economics and Statistics* 84:1 (2002), 116-130.

Saxenian, AnnaLee, *Silicon Valley's New Immigrant Entrepreneurs* (California: Public Policy Institute, 1999).

Schumacher, Dieter, "Market Size and Factor Endowment: Explaining Comparative Advantage in Bilateral Trade by Differences in Income and Per Capita Income," *Deutsches Institut für Wirtschaftsforschung Discussion Paper Series* No. 259 (2001).

SIL International, *Ethnologue: Languages of the World, 14th Edition* (www.ethnologue.com: Summer Institute of Linguistics, 2001).

Summers, Lawrence, "Regionalism and World Trading System," *Policy Implications of Trade and Currency Zones* (Wyoming: Proceedings of symposium sponsored by the Federal Reserve Bank of Kansas City, 1991).

Tinbergen, Jan, *Shaping the World Economy* (New York: The Twentieth Century Fund, 1962).

Venables, Anthony J., "Equilibrium Locations of Vertically Linked Industries," *International Economic Review* 37:2 (1996), 341-359.

Verbeek, Marno, *A Guide to Modern Econometrics* (Wiley: Chichester, 2000).

Witsa, *Digital Planet 2002: The Global Information Economy* (www.witsa.org: World Information Technology and Services Alliance, 2002).

World Bank, *World Development Indicators Database* (CD-rom: World Bank, 2001).

World Bank, *Little Data Book 2002* (Washington, D.C.: World Bank, 2002).

APPENDIX 1: CALCULATION OF THE DISTANCE VARIABLE

Definition Distance Variable

The distance variable is defined as the great circle distance between the capital of India (New Delhi) and the capitals of the various countries j.

Definition Alternative Distance variable

Great circle distance between the main cities (principal economic centres) of India (Mumbai) and the different countries j.

Based on Frankel (1997, p. 56, except for the choice of Mumbai instead of New Delhi).

Countries where the economic centre is not the capital:

- India:	Bombay (Mumbai)	instead of	New Delhi
- Australia:	Sidney	instead of	Canberra
- Bolivia:	La Paz	instead of	Sucre
- Brazil:	Sao Paulo	instead of	Brasilia
- China:	Shanghai	instead of	Beijing
- Morocco:	Casablanca	instead of	Rabat
- Nigeria:	Lagos	instead of	Abuja
- Pakistan:	Karachi	instead of	Islamabad
- Switzerland:	Geneva	instead of	Bern
- USA :	Chicago	instead of	Washington DC

Calculation:

Based on: Schumacher, Dieter (2001) 'Market size and Factor Endowment: Explaining Comparative Advantage in Bilateral Trade by Differences in Income and Per Capita Income', *Deutsches Institut für Wirtschaftsforschung Discussion Paper No. 259*, Berlin, Germany, August 2001, 21 pp.

Great circle distance d between two points with coordinates (lat_1, lon_1) and (lat_2, lon_2) in radians is given by:

$$D_{ij}(rad) = a\cos\left[\sin(lat_1) * \sin(lat_2) + \cos(lat_1) * \cos(lat_2) * \cos(lon_1 - lon_2)\right]$$

Conversion of the distance in radians to distance in nautical miles (nm):

$D_{ij}(nm) = [\ (180 * 60) / pi\ ] * D_{ij}(rad)$

APPENDIX 2: SIMPLE CORRELATION COEFFICIENTS BETWEEN INDEPENDENT VARIABLES (POOLED SOFTWARE SAMPLE: 1997-98-99-00)

Simple Correlation Coefficients between Independent Variables (Pooled Software Sample)					
	GDP_j	GDP_j/POP_j	D_{ij}	$LANG_j$	PIO_j
GDP_j	1	0.413853	0.193349	0.059588	-0.110600
GDP_j/POP_j	0.413853	1	-0.068741	-0.216356	-0.077500
D_{ij}	0.193349	-0.068741	1	0.137359	-0.243485
$LANG_j$	0.059588	-0.216356	0.137359	1	0.076051
PIO_j	-0.110600	-0.077500	-0.243485	0.076051	1

APPENDIX 3: SUMMARY STATISTICS FOR THE POOLED SOFTWARE SAMPLE (Excluding the Language Dummy)

Summary Statistics for the pooled Software sample (N = 213)					
Variable	X_{ij}	GDP_j	GDP_j/POP_j	D_{ij}	PIO_j
Mean	67.40	515,288.57	13,066.89	3,580.91	35,209.00
Median	8.01	125,887.40	11,122.82	3,118.94	1,240.28
Minimum	0.33	398.67	106.83	664.63	0.24
Maximum	3,857.59	9,837,406.00	38,161.77	9,130.34	603,432.98
Standard Deviation	337.31	1,362,908.79	11,046.74	1,934.65	98,935.96
Variance (in Millions)	0.11	1,857,520.38	122.03	3.74	9,788.32
Coefficient of Variation	5.00	2.64	0.85	0.54	2.81

APPENDIX 4: RESULTS OF ALTERNATIVE REGRESSIONS

Table 4.1: OLS Regression Results for the Pooled Sample (1997-98-99-2000) (Distance: between MUMBAI and the principal ECONOMIC CENTER of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX)		log(TOTEX)	
n	213		459	
Adj. R^2	0.637826		0.833466	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant-1997	-7.163722	-4.608326*	2.182555	2.963741*
Constant-1998	-7.001780	-4.517933*	2.445696	3.285937*
Constant-1999	-6.671352	-4.322362*	2.429160	3.289107*
Constant-2000	-6.505997	-4.202223*	2.588192	3.530446*
log(GDP)	0.492781	8.434305*	0.936393	38.74595*
log(GDP_j/POP_j)	0.301962	3.479554*	-0.260285	-7.003329*
log(Distance)	-0.053043	-0.261206	-0.860950	-9.242947*
Language	1.309472	5.034080*	0.132246	1.169434
log(PIO-share)	0.091212	2.161589*	0.161524	10.16811*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

Table 4.2: OLS Regression Results for the Pooled Sample (1997-98-99-2000): Alternative Specification (Distance: between Capitals) in which GDP_j is replaced by the population (or population related) variable				
	Software Exports		Total Exports	
Independent	log(SWEX)		log(TOTEX)	
n	210		459	
Adj. R^2	0.646538		0.821412	
Variables	Coefficient	T-statistic	Coefficient	T-statistic
Constant-1997	-5.656631	-3.412158*	-3.492269	-3.536098*
Constant-1998	-5.519415	-3.377231*	-3.229944	-3.260359*
Constant-1999	-5.233518	-3.213548*	-3.248312	-3.291342*
Constant-2000	-5.093869	-3.140348*	-3.084578	-3.147026*
log(GDP)	0.504508	9.036956*	0.634679	22.75518*
log($PCPOP_j$)	0.304744	4.559605*	-	-
log(POP_j)	-	-	0.301343	8.003434*
log(Distance)	-0.092570	-0.519811	-0.651817	-6.673910*
Language	1.197392	5.656339*	0.051008	0.427817
log(PIO-share)	0.103438	3.244010*	0.189529	11.26750*

* denotes that the coefficient is significant at least at the 90 % level in two-tailed test.

Table 4.3: Generalized Least Squares Regression Results for the Pooled Sample: Random Effects Model (Distance: between Delhi and the capital of country j)				
	Software Exports		Total Exports	
Independent	log(SWEX)		log(TOTEX)	
n	213		459	
Overall R^2	0.6225		0.8196	
Variables	Coefficient	Z-statistic	Coefficient	Z-statistic
Constant	-4.4560060	-2.04*	0.6266792	0.48
log(GDP)	0.4515470	5.23*	0.9301948	18.26*
log(GDP_j/POP_j)	0.3104102	2.30*	-0.3250037	-3.93*
log(Distance)	-0.2619000	-0.92	-0.5893233	-3.67*
Language	1.4622480	4.01*	0.0232354	0.10
log(PIO-share)	0.0391336	0.65	0.1932516	5.83*

* denotes that the coefficient is significant at least at the **95 % level** in two-tailed test.

APPENDIX 5: DATA SOURCES AND DESCRIPTION

The three major data sources pertaining to the Indian software are: the Department of Electronics (DoE), the Reserve Bank of India (RBI) and the National Association of Software and Service Companies (NASSCOM). The Centre for Monitoring the Indian Economy (CMIE) also gathers data on the Indian software traded in physical form (CDs, tapes, etc.) (For a critical appraisal of the contents and quality of the Indian software trade data, see Joseph and Harilal (2001))[17]. We have used the country-wise Indian software export data available from Nasscom. Note that the Nasscom data are compiled using firm level data supplied by its members who together account for 95 per cent of the revenue of the Indian software industry. The export figures reported by Nasscom include those originating from onsite services, offshore services and products and packages. The export values included in our sample account for 96.97%, 97.90%, 97.84% and 97.01% of the Indian exports of software for the years 1997-98, 1998-99, 1999-2000 and 2000-01 respectively[18]. For Total Exports the corresponding values are 94.37%, 92.92 %, 93.37% and 94.17% respectively. The Nasscom figures are reported in Indian rupees and we have converted them into U.S. Dollars using the yearly average exchange rates provided by the International Monetary Fund (IMF, International Financial Statistics, CD-rom (2001)).

India's bilateral total exports are from the Direction of Trade Statistics Yearbook (DOTSY) of the IMF (2001, pp. 256 – 258). IMF sources have confirmed to us that as software is not considered as merchandise trade, it is not included in the DOTS Series and hence is not a part of the total exports (communication by Ms. Olga Laveda, IMF, Washington D.C.). The GDP and GDP per capita figures are from the World Development Indicators Database (CD-rom, 2001) of the World Bank. The last three variables were available in U.S. Dollars. The POP_j and $PCPOP_j$ figures are also from the World Development Indicators Database (CD-rom, 2001). Some conceptual questions concerning the calculation of the distance variable were mentioned in section 3.B and the method used is explained in Appendix 1. In alternative calculations the distance between Mumbai (Bombay) and the main centers of economic activity (if considered to be different from the capital) of the different countries was used. The cities for the latter (with the exception of Mumbai) were identified on the basis of Frankel (1997, p. 56). The source for the latitudes and longitudes of the different cities is *The Getty Thesaurus of Geographic Names* of the Getty Research Institute. The

---

[17] Joseph and Harilal (2001) note that there is a difference in the recorded rate of growth of software exports in rupee and dollar terms due to an 'exchange rate effect'.

[18] Omitted countries include those for which data were grouped together as 'Others in Europe', 'Others in South East Asia', etc., and some individual countries for which matching explanatory variables could not be found. For some countries alternative sources were used to acquire data on GDP: for Taiwan, the source is Central Bank of China (2002); for Bahrain, Oman and United Arab Emirates: Little Data Book 2002 (2002); for Myanmar: International Financial Statistical Yearbook (IMF, 2001).

distance variable is expressed in nautical miles. The language variable takes the value of one in the case of countries (j's) where English is a national language. Such countries are identified from *Ethnologue: Languages of the World, 14th Edition* from SIL International. As was mentioned in subsection 3.B., the source of data on People of Indian Origin is the Report of the High Level Committee on the Indian Diaspora (Ministry of External Affairs, Government of India (2002)).