2014

English as a lingua franca in a preferential trading bloc

HAYLEY ROPPEL

THE UNIVERSITY OF AKRON

Abstract

The language barrier between countries has been recognized as an important transaction cost in the role of international trade. A common spoken language between two countries increases the volume of trade flows between them. Expanded on such research, Ku & Zussman (2010), Fidrmuc & Fidrmuc (2009), Hutchinson (2002), and Lee (2012) have found the English language to be trade facilitating. Previous research has lacked focus on the economic returns to a country's ability to speak English within a preferential trading bloc. While English may help to facilitate trade on a global scale, this paper finds that these findings also hold true for facilitating trade within an isolated/localized region. Within the Association of South-East Asian nations (ASEAN), a country's English language proficiency has a positive and significant effect on bilateral trade flows.

Contents

Introduction

English is the second most spoken language in the world and the first most spoken foreign language. It is currently the official language of 58 sovereign entities and 21 non-sovereign states. In comparison to the US dollar's role in international finance as a vehicle currency, English could play the role of a vehicle language (Ku & Zussman 2008). In order for two countries to communicate, whom have different official languages, English could facilitate communication cross borders between two countries. Expanded on previous research has been applications to the gravity model in the gains to trade from a country's ability to speak English , Ku & Zussman (2010), Fidrmuc & Fidrmuc (2009), Hutchinson (2002), and Lee (2012).

Previous research has lacked focus on the economic returns to a country's proficiency in English within a preferential trading bloc. While English may help to facilitate trade on a global scale, perhaps other languages facilitate trade better in a more isolated region. The Association of South-East Asian Nations, ASEAN, is comprised of 10 member nations: Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam. Within the ASEAN trade bloc, an array of official languages are spoken amongst these countries: Malay, English, Khmer, Indonesian, Lao, Chinese, Burmese, Filipino, Thai and Vietnamese, with not one country sharing the same official spoken language. Focusing on the ASEAN member countries, this paper intends to fill the gap in previous research about the importance of a country's English language proficiency on bilateral trade flows. Previous research has concluded that a country's proficiency in English tends to decrease transaction costs and increase the volume of trade occurring from an international perspective. I will test to see if the volume of trade flows increase between countries with higher levels of English proficiencies within the preferential trading bloc ASEAN. Of the ASEAN member countries, Malaysia,

Indonesia, Philippines, Thailand, Vietnam and Singapore are a part of the bamboo network[1], being heavily influenced by the Chinese culture. Due to the economic activity of Chinese businesses within these countries, it could be that Asian dialects have a larger effect than English in decreasing transaction costs, amongst the ASEAN countries.

Literature Review

The significance of a shared common language and its positive effects on bilateral trade flows in gravity models has been acknowledged by researchers before Frankel and Rose (2002), Anderson and van Wincoop (2004). Many approaches have been taken to analyze and incorporate the significance of communicating through such a non-native language, on bilateral trade flows. It is standard in most gravity models to include a dummy variable for the commonality of languages, but researchers have gone into further analysis to better measure the effects of language on bilateral trade flows.

Hutchinson (2002) examines the proportion of the population that speak English as a first language and the proportion that speak English as a second language. Hutchinson (2002) measures the percentage of English speakers based on the data provided by Crystal (1997). The data is run over three years, 1994-1996, and limited to 33 countries because lack of the trade data availability. A fixed effects model is run to correct for any biased results for single year data, resulting in little change in the results from the OLS regression. Hutchinson (2002) concludes that both countries with higher first language English speakers and second language English speakers will have higher levels of trade. In fact, Hutchinson (2002)'s results indicate English as a second language has a larger influence on trade for both imports and exports than English as a first language. Interestingly, English spoken as a second language increases trade

[1] The bamboo network consists of a network of Chinese owned businesses operating in Southeast Asia.

flows 1.3 times more for exports and 1.75 times more for imports compared to the proportion who speak English as a first language.

In a further analysis, Hutchinson (2005) studies international trade volumes, imports and exports disjointedly, using a linguistic distance measure. This measure is derived by using the linguistic scores of Chiswick and Miller (1998) which measures on a scale of 1 to 3 the difficultly of learning a foreign language, 1 being extremely difficult and 3 being least difficult. Hutchinson (2005) uses the inverse of this linguistic score and calls the inverse measure "linguistic distance". To correct for serial correlation from persistence effects, Hutchinson (2005) uses a Hatanaka 2-Step process and concludes that the higher a countries linguistic distance, the lower both imports and exports will be between two countries.

By measuring the mean TOEFL paper based test scores, from the Educational Testing Services' publication of TOEFL Test Score and Data Summary, covering 100 countries over the time period of 24 years from 1973 to 1997 Ku & Zussman (2010) create a proxy to measure English proficiency levels of a country. Using a 2SLS approach, Ku & Zussman (2010) proxy the English variable with what they call a "linguistic distance" measure obtained from data by Dyen et al. (1992) which studies phonetically how much two languages overlap. Dyen et al. (1992) data is limited to only Indo-European languages, restricting the sample size in Ku & Zussman (2010)'s regression. The results conclude little change from the OLS to the 2SLS regression and indicate English increases bilateral trade by 1.51 percent for every 1 percent increase in the product of TOEFL scores.

The Eurobarameter survey was used by Fidrmuc & Fidrmuc (2009) and Lee (2012) to measure English proficiencies of 29 European Union countries in 32 languages. In the survey, which was restricted to only citizens of the European Union, participants were asked their

mother tongue and up to three other languages they could successfully communicate in. This measure improves Meltiz (2000) measure of language because it takes into account foreign languages instead of simply indigenous languages. First using the traditional dummies for English, French, German, Swedish, Dutch and Greek, Fidrmuc & Fidrmuc (2009) go on to use a 2SLS approach controlling for linguistic similarities between countries. Focused within the European Union, Fidrmuc & Fidrmuc (2009) find English is indeed an important role in facilitating trade, restricting such findings to that of Western Europe. Their models indicate that "an increase in English proficiency in all EU15 countries by 10 percentage points (keeping UK and Irish proficiency levels constant) would increase the intra-EU15 trade by 15% on average" Fidrmuc & Fidrmuc (2009).

Similar to Fidrmuc & Fidrmuc (2009), Lee (2012) uses the Eurobarameter survey (2006) although in a different retrospect, focusing on international trade in services. Lee (2012) expands the sample to non-European countries, which are not within the Eurobarameter survey, using governmental official homepage data taken from Mo (2010). In this aspect, Lee (2012) applies English proficiency levels to trade in services instead of overall bilateral trade flows by nine different sub-sectors: communication, computer and information, finance, insurance, business, commerce, royalties and license fees, transportation and travel among 22 OECD countries. Lee (2012) uses the Poisson pseudo-maximum likelihood estimation method, and finds that higher English language proficiencies have positive and significant importance on trade in services, particularly the communication, finance, commerce, insurance and business sectors.

Theoretical Model

The model used to control the determinants of bilateral trade flows is the gravity model:

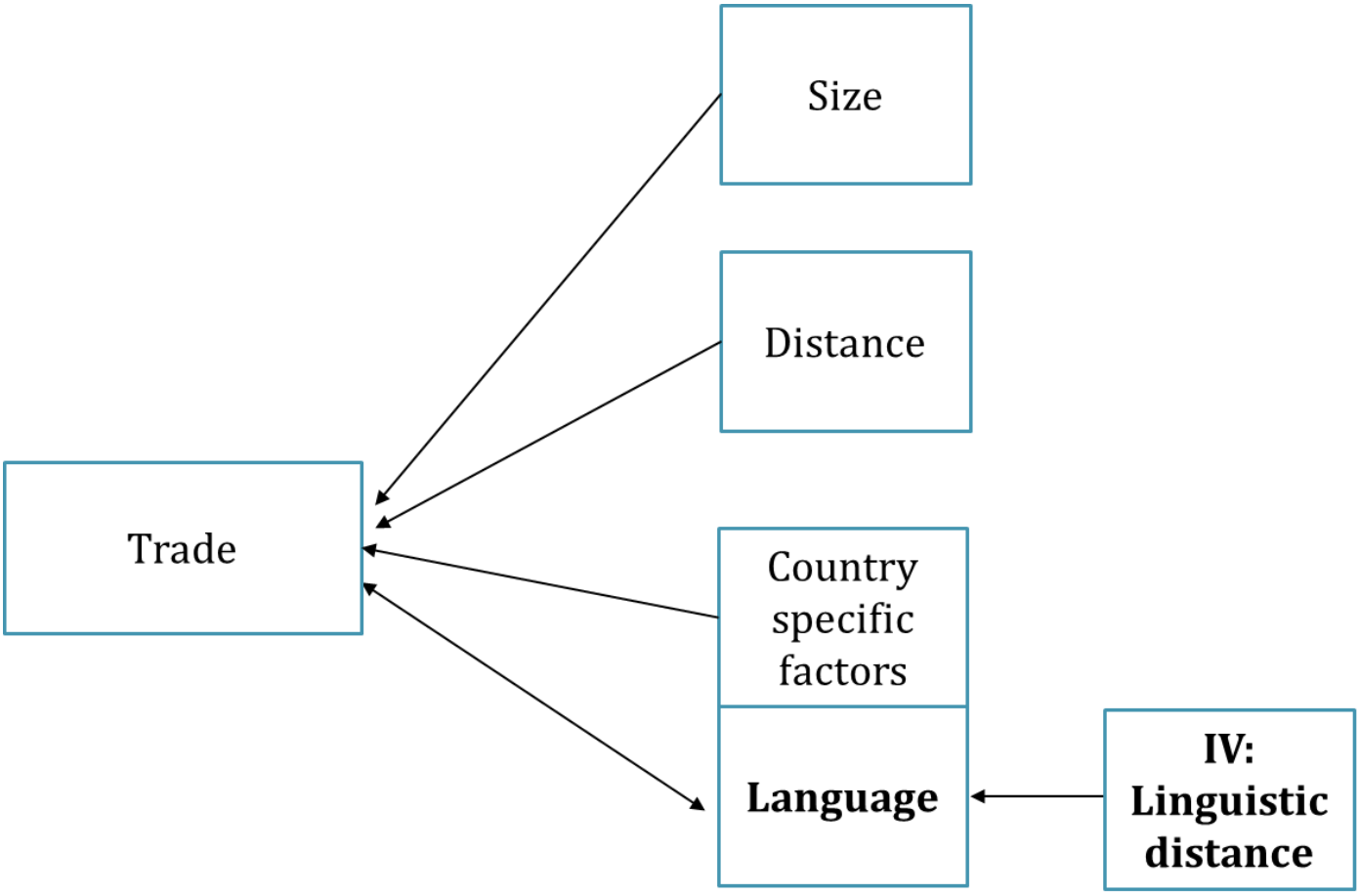
$$trade_{ij} = A \frac{GDP_i * GDP_j}{Distance_i * Distance_j}$$

Where, i and j represent two trading countries, in which the baseline volume of trade flows between the pair is directly related to the size of each country and indirectly related to the distances between the pair of countries. Other country specific factors, A, can have positive or negative effects on trade between two country pairs. In logarithmic form, the gravity model becomes:

$$trade_{ij} = \beta_0 + \beta_1 lnGDP_i + \beta_2 lnGDP_j + \beta_3 lnDistance_{ij}$$

Where GDP is used to control for the size of each country, in which β_1 and β_2 are expected to have positive coefficients, as larger countries can produce and consume more and β_3 is expected to have a negative coefficient, as distances increase this results in higher transportation costs. In Figure I, a further explanation of the gravity model is explained including the variable of interest for this study, language. Of the country specific factors effecting trade, language is included, expected to positively affect trade flows due to decreasing the transaction costs between two countries. There is a problem in that language can be spread as a result of increased bilateral trade flows, therefore, arising an endogeneity issue between trade and language. This must be controlled for by an instrumental variable that is not affected by increased trade over time, such as a language's linguistic distance. In this way, the model will be less biased and better predict the true effects of language on bilateral trade flows.

Figure I


Size
Distance
Trade
Country specific factors
Language
IV: Linguistic distance

Empirical Model

This paper pulls from the theoretical framework used by Fidrmuc & Fidrmuc (2009) and Ku & Zussman (2010). The empirical model used is similar to their models adjusted to account for a preferential trading bloc, the cultural past of ASEAN countries, and re-identify the common language variable due to each country having a different official language. The specification of the gravity model that I will be using is the following:

$$TRADE_{ij} = \beta_0 + \beta_1 lnGDP_i + \beta_2 lnDIST_j + \beta_3 lnENGPROF_{ij} + \beta_4 LANDLOC + \beta_5 ISLAND$$
$$\beta_6 FCOL + \beta_7 BORDER + \beta_8 CHINESE_{ij} + \varepsilon ,$$

Where i and j represent trading partners and LANDLOC, ISLAND, FCOL, BORDER, and COMLANG are dummy variables.

Data

The data are for the year 2012 including all ASEAN countries, with the exception of Brunei Darussalam[2]. In order to measure a country's proficiency in English, $ENGPROF_{ij}$, the mean score of examinees from each country in the Test of English as a Foreign Language (TOEFL) are used. The test data is managed by the Educational Testing Services (ETS) and is a standardized test for non-native speakers of English. TOEFL scores are used by more than 9,000 universities in 130 countries for enrollment in US universities and English speaking institutions. $ENGPROF_{ij}$ is calculated by multiplying two trade partners' mean TOEFL scores. For the purpose of this paper, I use only the 2012 TOEFL Test Score and Data Summary scores which are computer based tests. Prior to 2012 TOEFL exams were paper-based and not reported annually. TOEFL scores are scored out of 100 points across four sections: Reading, Writing,

[2] TOEFL score data for Brunei unavailable.

Speaking, and Listening. For the ASEAN countries, scores ranged from a low of 68 to a high of 98.

Other country specific variables included that are expected to positively affect bilateral trade flows are: the gross domestic product of $country_i$ and $country_j$, GDP_i and GDP_j, measured in current US dollars; if the a country in the pair was colonized by the French, FCOL; and if the two countries share a land border, BORDER. The variables expected to negatively affect trade flows are: the distance between $country_i$ and $country_j$, $DIST_{ij}$, defined as the great circle distance[3]; if a country in the pair is landlocked, LANDLOC; and if a country in the pair is an island, ISLAND. Also included is a dummy, $CHINESE_{ij}$, which contains a value of 1 if the Chinese language is recognized as being spoken in both trading countries. Table I contains the descriptive statistics for all trade variables.

Empirical Results

The results of the Ordinary Least Square regression are displayed in column (1) of Table II. The model is run over 56 observations and explains 83.2 percent of the variation in the data. The intercept, GDP variables, distance, English proficiency, and dummies for being a landlocked country, colonized by the French, and sharing a land border are statistically significant and have their expected signs. The variable capturing if Chinese is a recognized language spoken in both trading partner countries is positive, but it is not significant. The variable of interest $ENGPROF_{ij}$, shows a positive and significant effect on trade flows between ASEAN countries. A 1 percentage increase in the mean product of TOEFL scores between $country_i$ and $country_j$ should increase bilateral trade flows by 4.79 percent.

[3] Air distance between each trading pairs' capitals

This OLS model has treated the English proficiency variable as an exogenous variable, when in fact it is an endogenous variable. Not only is it that the trade flows between two countries are affected by levels of English proficiency, but it could be that English proficiency levels increase as a result of increased trade flows. In order to control for this issue of endogeneity between trade flows and English proficiency, causing biased results in this regression, my next step will be to run a Two-Stage Least Squares regression (2SLS). By running a 2SLS I will be able to control for $ENGPROF_{ij}$ as an endogenous variable by using an instrumental variable that is not effected by trade flows.

2SLS

Using the data from Chiswick and Miller (1998), the linguistic distance measure is used as a proxy for English Proficiency. This data is derived from matching each country code from the 1990 Census language codes to the linguistic scores (LS) and then using the inverse (1/LS) as the linguistic distance (LD) measure. The 2SLS regression results are displayed in column (2) of Table II. The R-squared is slightly decreased explaining 80 percent of the variation in the data now. The results agree English proficiency to be positively and significantly affecting trade flows and suggest that the OLS estimates understate the language impact. Controlled for endogeneity, the results predict that increasing the English language proficiency between trading partners by 1 percent should increase bilateral trade flows by 6.97 percent. The distance variable becomes insignificant in the 2SLS regression. This may be the result of the countries being already geographically close to each other therefore, the distance not having much impact on trade decisions. The Chinese dummy variable increases, yet remains insignificant in the 2SLS.

Conclusion

The results on the effects of the English proficiency of a country on bilateral trade flows is in line with the findings of previous research. While English language proficiency is trade facilitating on an international level, this is also true for facilitating trade within the ASEAN preferential trading bloc. After accounting for the endogeneity issue between the movement of language and trade volumes, the 2SLS regression results for English proficiency are positive and significant at the 90 percent confidence level. Increasing the English proficiency between two countries in the ASEAN trading bloc by 1 percent should result in increased trade flows between them by about 7 percent. This is a substantial impact affecting bilateral trade volumes. Including a dummy to capture the effect of the Chinese language within this trading bloc remained insignificant in both regressions, implying that English has a greater influence on trade flows despite the Chinese cultural influence in these countries.

Limitations to this study include the lack of data available in that there were no TOEFL data reported for Brunei Darussalam and the TOEFL data was only available for the year of 2012 due to changes in the test format and that prior to 2012 TOEFL data were not reported annually. Other data limitations include the lack of data to capture Chinese language proficiencies in the ASEAN countries, in which resulted in a dummy for if the language was recognized at all in a country regardless of the proficiency or amount of speakers. Future research can continue to verify that the results are sound by running this regression over more years once new data are published. Also, a better defining of the Chinese language within these countries, such as a Chinese language proficiency index, could better describe the impact of the Chinese culture within ASEAN countries.

Table I

Variable	Mean	Std. Dev.	Source
$lnTRADE_{ij}$	10358897583	17229628464	UN Comtrade
$lnGDP_i$	26.04	1.21	WorldBank
$lnGDP_j$	25.49	1.49	WorldBank
$lnDIST_{ij}$	7.27	0.57	TimeandDate
$lnENGPROF_{ij}$	8.79	0.15	ETS
LANDLOC	0.13	0.33	
ISLAND	0.79	0.41	
FCOL	0.55	0.50	
BORDER	0.21	0.41	
$CHINESE_{ij}$	0.11	0.50	CIA Factbook
$lnLINGDIST_{ij}$	-1.44	0.29	Chiswick & Miller (1998)

Note: The data include observations for 2012. Brunei Darussalam is excluded due to no TOEFL score data available.

Table II

	(1)	(2)
Variable	**OLS**	**2SLS**
Intercept	**-81.31*** (-3.73)**	**-111.42*** (-2.75)**
$\ln GDP_i$	**1.44*** (7.67)**	**1.60*** (7.24)**
$\ln GDP_j$	**1.25*** (6.79)**	**1.38*** (6.55)**
$\ln DIST_{ij}$	**-1.25*** (-2.91)**	**-0.69 (-1.12)**
$\ln ENGPROF_{ij}$	**4.78** (2.11)**	**6.97* (1.84)**
Landloc	**-1.37** (-1.97)**	**-0.74 (-0.89)**
Island	**-0.67 (-0.96)**	**-1.49 (-1.45)**
Fcol	**1.04* (1.69)**	**0.98 (1.46)**
Border	**1.91*** (3.15)**	**1.54** (2.47)**
Chinese	**0.57 (0.81)**	**0.99 (1.40)**
R^2	**0.8320**	**0.8029**
Adj. R^2	**0.7991**	**0.7644**

Note: t-stats in (_); N = 56

*Significance levels: *** 99 percent, ** 95 percent, * 90 percent*

Bibliography

Chiswick, B., & Miller, P. (). English language fluency amoung immigrants in the United States. *Research in labor economics*, *17*, 151-200

Fidrmuc, J., & Fidrmuc, J. (). Foreign Languages and Trade. *Brunel University West London*, *09*, 1-29.

Hutchinson, W. (). Does ease of communication increase trade? Commonality of language and bilateral trade. *Vanderbilt University*, *02*, 1-25.

Hutchinson, W. (). "Linguistic distance" as a determinant of bilateral trade. *Southern Economic Journal*, *72*, 1-15.

Ku, H., & Zussman, A. (). Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization*, *75*, 250-260.

Lee, K. (). International trade in services and the role of English. *Journal of East Asian Economic Integration*, *16*, 291-314.

SAS coding:

```
LIBNAME SP "E:\SP\data";

PROC IMPORT DATAFILE = "E:\SP\data\Datanew2.csv"
      DBMS = CSV
      OUT = SP.TRADE
      REPLACE;
RUN;

data trade1;
      set sp.trade;

            IMEX= (IM+EX);
            AREA = Areai * Areaj;
            EngProf = EngProfi * EngProfj;
            lnGDPi= log(GDPi);
            lnGDPj= log(GDPj);
            lnDIST= log(DISTij);
run;

data trade2;
      set trade1;

            lnEngProf = log(EngProf);
            lnTRADE= log(IMEX);

      run;
proc reg;

model lnTRADE = lnGDPi lnGDPj lnDIST lnEngProf landloc island fcol Border
Chinese;
run;


data trade3;
      set trade2;
      IV=lingdisti*lingdistj;
      run;
data trade4;
      set trade3;
      LD=(1/IV);
      run;
data trade5;
      set trade4;
      lnLingDIST= log(LD);
      run;

proc syslin 2SLS;
endogenous lnEngProf;
instruments lnLingDIST  ;
first_stage: model lnEngProf= lnLingDIST ;
second_stage: model lnTRADE = lnGDPi lnGDPj lnDIST lnEngProf landloc island
fcol Border Chinese;
run;
```