The Effect of Knowledge Accessibility on International Income Inequality*

Kentaro Nakajima†

February 21, 2006

Abstract

Using a structural model of new economic geography, we estimate the effect of accessibility to foreign knowledge on international income distribution. Whereas previous literature has mainly focused on the importance of geographical accessibility to foreign markets, this paper emphasizes the role of accessibility to foreign knowledge in determining international income inequality. Using cross-country data of income per worker, bilateral trade flows, and number of flight passengers, we find evidence that the accessibility to foreign knowledge raises per capita income significantly.

Keywords: New economic geography; Market potential; Gravity equation; Knowledge spillovers

JEL classification: F12; O31; R12

*I am very grateful to Takatoshi Tabuchi for his encouragement, helpful comments and careful reading of this paper. I am also grateful to two anonymous referees for useful comments and suggestions. I also thank Kristian Behrens, Masahisa Fujita, Hidehiko Ichimura, Hiro Ishise, Yoshitsugu Kanemoto, Tomoya Mori, Komei Sasaki, Yasuyuki Sawada, Tetsushi Sonobe, Kazuhiro Yamamoto and seminar participants at the Microeconomics Workshop, Empirical Micro Brown-Bag Lunch Seminar at the University of Tokyo, Japanese Economic Association in Tokyo, and Applied Regional Science Conference in Chiba for their useful comments and suggestions. I also thank Roger Smith for careful reading of this paper. Of course, any remaining errors are my own.

†Graduate School of Economics, the University of Tokyo, `kentaro@grad.e.u-tokyo.ac.jp`

1 Introduction

International income inequality is one of the most important problems in the world economy today. According to the World Development Report (World Bank, 1995), 1.4 billion of 2.5 billion workers worldwide live in poor countries (which are defined as those countries with annual income per capita below \$695) whereas some 380 million workers live in high income countries and can earn an annual income above \$8626. While there are many factors that cause unequal income distribution, we focus on geographical factors by investigating their effect empirically using a structural model of *new economic geography* (NEG). NEG explains spatial aspects of the economy through the interaction between transport cost and increasing returns to scale, which stems from the research of Krugman (1980, 1991).

Recently, several empirical studies have investigated the geographical effect on international income disparity by using structural NEG models (Head and Mayer, 2004a). Such researches have investigated the effect of *market potential*, which indicates the geographical aspects of demand, on the manufacturing wage. Hanson (2005) examined cross-region variations in the United States, while Redding and Venables (2004) investigated cross-country relationships. Both studies found evidence that market potential positively and significantly affects the manufacturing wages.

In contrast to such empirical works which mainly focused on market potential, there are several theoretical NEG studies emphasizing knowledge spillovers as well as market potential (Baldwin and Martin, 2004). Baldwin et al. (2001) have incorporated knowledge spillovers into a standard NEG model using the endogenous growth model of Grossman and Helpman (1991) showing that foreign knowledge spillovers as well as pecuniary externalities affect the spatial equilibrium. Their results imply the possibility that the accessibility to foreign knowledge, as well as market potential, affect the manufacturing wage.

Several empirical studies have investigated the effect of foreign knowledge spillovers on the economy and confirmed its importance. For example, Coe and Helpman (1995) used trade flows as a proxy of knowledge flows and investigated the positive relationship between knowledge spillover and productivity. Peri (2005) used patent citations as a proxy of knowledge flows, and investigated the relationship between knowledge spillover and innovation. Keller (2002) estimated the effect of foreign R&D spillovers on productivity. The validity of Baldwin et al. (2001) conjecture, relationship between knowledge spillover and manufacturing wage, however, has not been tested empirically to the best of our knowledge.

In this paper, we empirically test the effect of accessibility to foreign knowledge as well as market potential on manufacturing wage by using a structural NEG model. We extend the model of Redding and Venables (2004) using the empirical models of knowledge spillovers (Keller, 2002, Peri, 2005). We use cross-country data of income per worker, bilateral trade flows, and the number of flight passengers as a proxy of knowledge flows. Our results suggest that accessibility to foreign knowledge positively and significantly affects the manufacturing wage. This supports the approach taken by Baldwin et al. (2001) in showing that knowledge spillovers affect spatial equilibrium.

This paper is organized as follows. Section 2 presents a theoretical NEG framework, and defines *knowledge access* as the geographical accessibility to foreign knowledge. Section 3 obtains the estimator of knowledge access. Section 4 estimates the wage equation and shows the significant effect of knowledge access on income variations. Section 5 offers some conclusions.

2 Theoretical framework

In this section, we present an NEG model based on Redding and Venables (2004) with the introduction of intermediate goods and a knowledge spillovers structure. In the next subsection, we provide the basis of the NEG framework, and in subsection 2.2, we present the knowledge spillover structure.

2.1 NEG model

The world consists of $i \in \{1, \ldots, R\}$ countries. Our model is focused only on the manufacturing sector which produces differentiated goods and which are used both in consumption and as intermediate goods.

We start the analysis on the demand side. Each firm's product is differentiated from that of other firms and is used both in consumption and as an intermediate good. We assume both utility and production functions are constant elasticity of substitution (CES) with σ representing an index of product differentiation,

$$U_j = \left[\sum_{i=1}^{R} \int_0^{n_i} x_{ij}(z)^{(\sigma-1)/\sigma} dz\right]^{\sigma/(\sigma-1)}, \quad \sigma > 1,$$

where z is manufacturing varieties, n_i is the set of varieties produced in country i, and $x_{ij}(z)$ is the demand of country j for the z-th product produced in country i. Solving the maximization problem, we obtain country j's total demand for each good

$$x_{ij} = p_{ij}^{-\sigma} E_j G_j^{\sigma-1}, \tag{1}$$

where p_{ij} is the prices of each variety produced in country i and sold in j, G_j is the price index of country j, $G_j = \left[\sum_i \int_0^{n_i} p_{ij}(z)^{1-\sigma} dz\right]^{1/(1-\sigma)}$, and E_j is country j's total expenditure on manufacturing goods which is used as consumption and intermediate goods. Therefore, the term $E_j G_j^{\sigma-1}$, which is the manufacturing expenditure adjusted by the price index, can be interpreted as the real expenditure on manufacturing goods.

Next, we turn to the analysis of the production side. We assume that a firm's profit function in country i is

$$\pi_i = \sum_j p_{ij} x_{ij}/T_{ij} - G_i^{\alpha} w_i^{\beta} v_i^{\gamma} (1/h_i)(F + x_i), \tag{2}$$

where $x_i \equiv \sum_j x_{ij}$. This technology exhibits increasing returns to scale with $F > 0$. There are three types of inputs; intermediate goods with the price index G_i^{α} and share α; labor with the wage w_i and share β; and mobile factor with price v_i and share γ. Assume that $\alpha + \beta + \gamma = 1$, and h_i is the total factor productivity (TFP) of each firm located in country i. Thus, $1/h_i$ can be interpreted as the production efficiency of country i. T_{ij} is the iceberg transport cost between country i and j. If one unit of the manufacturing good shipped from country i to j, only a fraction $1/T_{ij}$ of original unit actually arrives. If trade is costless, $T_{ij} = 1$.

Maximizing this profit function, we have

$$p_i = G_i^{\alpha} w_i^{\beta} v_i^{\gamma} (1/h_i)\sigma/(\sigma - 1). \tag{3}$$

where p_i holds $p_{ij} = p_i T_{ij}$. We suppose that there is free entry and exit in response to profits or losses, so that firm's zero-profit conditions are satisfied, and using the demand function eq. (1), we have

$$p_i^{\sigma}(\sigma - 1)F = \sum_j E_j G_j^{\sigma-1} (T_{ij})^{1-\sigma}. \tag{4}$$

Substituting eq. (3) into eq. (4) yields a so-called wage equation;

$$(\sigma - 1)F[G_i^{\alpha} w_i^{\beta} v_i^{\gamma} (1/h_i)\sigma/(\sigma - 1)]^{\sigma} = \sum_j E_j G_j^{\sigma-1} T_{ij}^{1-\sigma}. \tag{5}$$

This wage equation represents the price of the immobile factor of production.

Next, we define trade equation to analyze market access and supplier access. Multiplying both sides of eq. (1) by $n_i p_i$, we obtain the trade equation

$$n_i p_i x_{ij} = n_i p_i^{1-\sigma} (T_{ij})^{1-\sigma} E_j G_j^{\sigma-1}. \tag{6}$$

The left-hand side denotes the aggregate value of exports. The right-hand side of this equation contains both demand and supply variables: $E_j G_j^{\sigma-1}$ is country j's market capacity, while $n_i p_i^{1-\sigma}$ is the supply capacity. They are argued in detail in the next section. For the sake of simplicity, we assume E_j and n_j are exogenous.

Next we define market access[1] (MA) and supplier access (SA). Market access is the transport cost weighted sum of the market capacities ($E_i G_i^{\sigma-1}$) of all partner countries , and supplier access is that of supplier capacities ($n_i p_i^{1-\sigma}$); therefore:

$$\mathrm{MA}_i = \sum_j E_j G_j^{\sigma-1} (T_{ij})^{1-\sigma}, \quad \mathrm{SA}_j = \sum_i n_i (p_i T_{ij})^{1-\sigma}. \tag{7}$$

These are the definitions of market and supplier access. Substituting eq. (7) into eq. (5), we can

[1]This is the nominal market potential defined by Head and Mayer (2004a). Redding and Venables (2004) have called it market access. Since our model largely depends on Redding and Venables (2004), we follow their expression.

rewrite the wage equation as

$$[w_i^{\beta} v_i^{\gamma} (1/h_i)]^{\sigma} = A G_i^{-\alpha\sigma} \sum_j E_j G_j^{\sigma-1} (T_{ij})^{1-\sigma} = A \left(\sum_j n_j p_j^{1-\sigma} (T_{ij})^{1-\sigma} \right)^{\frac{\alpha\sigma}{\sigma-1}} \left(\sum_j (T_{ij})^{1-\sigma} E_j G_j^{\sigma-1} \right)$$
$$= A(\mathrm{SA}_i)^{\frac{\alpha\sigma}{\sigma-1}} (\mathrm{MA}_i), \tag{8}$$

where A is a bundle of parameters.

2.2 Knowledge access

In this subsection, we extend the wage equation by incorporating knowledge spillovers. First, we define the structure of the TFP. Based on Keller (2002) and Peri (2005), we assume the TFP of a firm located in country i is determined by knowledge access (KA_i), $h_i = \mathrm{KA}_i$, and knowledge access is defined by the sum of the accessible ideas:

$$\mathrm{KA}_i = \sum_j^R \phi_{ij} \Upsilon_j.$$

where Υ_j is the ideas generated in country j and $\phi_{ij} \in [0, 1]$ is the fraction of ideas generated in country j that is accessible to country i. Then, the wage equation (8) is modified as follows

$$(w_i^{\beta} v_i^{\gamma} \mathrm{KA}_i)^{\sigma} = A(\mathrm{SA}_i)^{\frac{\alpha\sigma}{\sigma-1}} (\mathrm{MA}_i). \tag{9}$$

This is the wage equation we want to estimate in this paper.

Next, following Peri's (2005) formula, we specify ϕ_{ij} as

$$\phi_{ij} = (\mathrm{dist}_{ij})^{\delta_1^K} \exp[\delta_2^K (\mathrm{bord}_{ij}) + \delta_3^K (\mathrm{lang}_{ij}) + \delta_4^K (\mathrm{trbl}_{ij})]. \tag{10}$$

There are four geographic components in eq. (10): dist_{ij} is equal to $\mathrm{dist}_{ii} = (2/3)\sqrt{\mathrm{area}_i/\pi}$ if $i = j$, where area_i denotes the area of the country i (otherwise the distance between capitals of countries i and j)[2]; bord_{ij} is a dummy which equals 1 if countries i and j share a border

[2] This formula is followed from Redding and Venables (2004) and Head and Mayer (2004b).

and 0 otherwise; lang_{ij} is equal to 1 if the same language is spoken in countries i and j and 0 otherwise; and trbl_{ij} is 1 if countries i and j belong to the same trade bloc and 0 otherwise.

Actual flows of knowledge from countries j to i (c_{ij}) are assumed as:

$$c_{ij} = e^{\mu_i^K} \phi_{ij} \Upsilon_j e^{\varepsilon_{ij}}, \tag{11}$$

where μ_i^K is the fixed effect of receiving country i and ε_{ij} is the stochastic disturbance.

3 First stage estimation

Following Redding and Venables (2004), we estimate the wage equation by a two-stage procedure. In the first stage, we estimate the trade equation and construct the fitted value of the market and supplier access. In the second stage, using the fitted value of market and supplier access, we estimate the wage equation.

3.1 Data sources and sample size

Data on the bilateral trade flows for between countries are obtained from the World Bank's COMTRADE database. We use the number of flight passengers between country's primary airports as a proxy of knowledge flows, because face-to-face communications is the most important knowledge spillover. Using primary airports, we can remove a certain amount of tourists. However, we cannot remove these factors completely. In order to check validity of the data, we also use the amount of bilateral mail flows between primary airports as another proxy for knowledge flows. Data on the bilateral flight passengers and the mail flows are obtained for a cross-section from ICAO's On-Flight Origin and Destination.

Because there are many missing values and large variations in these data, especially in the number of the flight passengers, we take the average between 1995 to 1999, which smooths the year-specific effects.

3.2 Trade equation estimation

In this subsection, we estimate the trade equation and construct market access and supplier access. We specify transport cost $T_{ij}^{1-\sigma}$ as below

$$T_{ij}^{1-\sigma} = (\text{dist}_{ij})^{\delta_1^T} \exp[\delta_2^T(\text{bord}_{ij}) + \delta_3^T(\text{lang}_{ij}) + \delta_4^T(\text{trbl}_{ij})]$$

This specification is the same as the knowledge spillovers fraction ϕ_{ij} in eq. (10). Substituting this $T_{ij}^{1-\sigma}$ specification into trade equation (6), and taking logarithms, we obtain the estimable specification as follows:

$$\ln(\text{EX}_{ij}) = \theta^T + \mu_i^T \text{cty}_i + \lambda_j^T \text{ptn}_j + \delta_1^T \ln(\text{dist}_{ij}) + \delta_2^T \text{bord}_{ij} + \delta_3^T \text{lang}_{ij} + \delta_4^T \text{trbl}_{ij} + u_{ij}^T. \quad (12)$$

where EX_{ij} is the value of exports from countries i to j; cty_i and ptn_j is the dummy variable of country i and partner j, which capture supply and market capacities[3].

Using estimated values $\hat{\mu}_i^T, \hat{\lambda}_j^T, \hat{\delta}_1^T$, and $\hat{\delta}_2^T$, we construct $\widehat{\text{MA}}_i$ and $\widehat{\text{SA}}_j$ from eq. (7) as below

$$\widehat{\text{MA}}_i = \text{D}\widehat{\text{MA}}_i + \text{F}\widehat{\text{MA}}_i = (\exp(\text{ptn}_i))^{\hat{\lambda}_i^T}(T_{ii})^{1-\sigma} + \sum_{j\neq i}(\exp(\text{ptn}_j))^{\hat{\lambda}_j^T}(T_{ij})^{1-\sigma},$$

$$\widehat{\text{SA}}_j = \text{D}\widehat{\text{SA}}_j + \text{F}\widehat{\text{SA}}_j = (\exp(\text{cty}_j))^{\hat{\mu}_j^T}(T_{jj})^{1-\sigma} + \sum_{i\neq j}(\exp(\text{cty}_i))^{\hat{\mu}_j^T}(T_{ij})^{1-\sigma},$$

where D represents the domestic effect and F the foreign effect.

Using ordinary least squares (OLS), we estimate the trade equation eq. (12) with results shown in Table 1 . The coefficient of ln(dist) is negative, and the coefficient of common border

Table 1: Trade equation estimation

Dependent	Observations	Year	lndist	bord	lang	trbl	cty	ptn	R^2	F(·)	Prob>F
ln(EX)	2931	95–99	−0.945**	0.731**	1.237**	0.335**	yes	yes	0.854	134.7	0.000
			(−21.74)	(4.68)	(13.42)	(3.11)					

Heteroskedasticity-robust t-values in parentheses.
*denotes statistical significance at the 10% level. **denotes statistical significance at the 5% level.

[3]Since we cannot observe completely these economic variables that correspond exactly to the theory, we use these dummies as a substitute following Redding and Venables(2004).

and language are positive. Both coefficients are statistically significant at the 1% level, thus vindicating the model. The null hypothesis, the coefficients of the country dummies, and partner dummies equal to zero are rejected at the 1% level by an F-test. Then we construct $\widehat{\mathrm{MA}}_i$ and $\widehat{\mathrm{SA}}_j$ using these estimated coefficients.

3.3 Knowledge access estimation

In this subsection, we construct knowledge access and include it in the estimation. Substituting eq. (10) into eq. (11), and taking the logarithm of it, we obtain the estimation function as shown below,

$$\ln(c_{ij}) = \theta^K + \mu_i^K \mathrm{cty}_i + \lambda_j^K \mathrm{ptn}_j + \delta_1^K \ln(\mathrm{dist}_{ij}) + \delta_2^K \mathrm{bord}_{ij} + \delta_3^K \mathrm{lang}_{ij} + \delta_4^K \mathrm{trbl}_{ij} + u_{ij}^K, \tag{13}$$

where $\lambda_j \equiv \ln \Upsilon_j$. We use partner dummies as the proxy of the ideas generated in country j, similar to the trade equation estimation, because we cannot obtain the data of generated ideas.

Substituting estimated parameters $\hat{\mu}_i^K, \hat{\lambda}_j^K, \hat{\delta}_1^K$,and $\hat{\delta}_2^K$ into the definition of knowledge access eq.(9), we construct a fitted value of knowledge access $\widehat{\mathrm{KA}}_i$ as

$$\widehat{\mathrm{KA}}_i = \widehat{\mathrm{DKA}}_i + \widehat{\mathrm{FKA}}_i = (\exp(\mathrm{ptn}_i))^{\hat{\lambda}_i^K} \phi_{ii} + \sum_{j \neq i} (\exp(\mathrm{ptn}_j))^{\hat{\lambda}_j^K} \phi_{ij}.$$

As with the trade equation estimation, we regress eq. (13) by OLS. We use the number of flight passengers from country j to i ,TR_{ij}, as a proxy of knowledge flows from j to i because we consider face-to-face communications are the most important knowledge spillovers[4]. However, in order to check validity of the data, we also use another proxy of knowledge flows. We use the amount of mail flows from country j to i, MAIL_{ij}, as another proxy of knowledge flows from j to i. The estimation result is shown in Table 2 . Column (1) uses the number of flight passengers as the flows of knowledge ($c_{ij} = \mathrm{TR}_{ij}$). Similar to the result of the trade equation estimation, the coefficient of log distance is negative, those of common border dummy and common language

[4]Even within a city, face-to-face communications are very important as is often discussed in urban economics. See Fujita and Thisse (2002), among others.

Table 2: Knowledge access estimation

	Dependent	Observations	Year	ln(dist)	bord	lang	trbl	cty	ptn	R^2	F(·)	Prob>F
(1)	ln(TR)	894	95–99	−0.678**	0.379**	0.622**	0.079	yes	yes	0.674	20.35	0.000
				(−9.78)	(3.83)	(4.79)	(0.51)					
(2)	ln(MAIL)	850	95–99	−0.574**	0.146	1.019**	0.134	yes	yes	0.642	23.37	0.000
				(−5.18)	(0.81)	(5.32)	(0.53)					

Heteroskedasticity-robust *t*-values in parentheses.
*denotes statistical significance at the 10% level. **denotes statistical significance at the 5% level.

dummy are positive, which supports our model. These coefficients are statistically significant at the 1% level. On the other hand, the coefficient of trade bloc is not significant at the 10 % level. Column (2) uses the amount of mails ($c_{ij} = \mathrm{MAIL}_{ij}$). In the results, the coefficient of log distance is also negative and common language dummy are positive. In both estimation, the null hypothesis of zero country dummies and zero partner dummies are rejected by the F-test.

The results of column (2) are similar enough to that of column (1), it implies that the existence of tourists does not seem to bias the estimation using flight passenger data. Therefore, we construct knowledge access using column (1) results and use it in the main estimations of the wage equation.

4 Estimation of wage equation

We now estimate the wage equation using knowledge access, market access, and supplier access as obtained by the first stage estimation.

4.1 Specification and data

Rearranging the wage equation eq. (9) by taking logarithms, we can derive an estimation function

$$\ln w_i = \xi + \psi_1 \ln \mathrm{KA}_i + \psi_2 \ln \mathrm{MA}_i + \psi_3 \ln \mathrm{SA}_i + \eta_i.$$

Substituting the fitted value of knowledge, market, and supplier access, we obtain the following estimation function:

$$\ln w_i = \zeta + \psi_1 \ln \widehat{\mathrm{KA}_i} + \psi_2 \ln \widehat{\mathrm{MA}_i} + \psi_3 \ln \widehat{\mathrm{SA}_i} + \varepsilon_i. \tag{14}$$

GDP per worker from World Development Indicator is used as the manufacturing wage w_i, and we take the average of 1995-1999. $\widehat{\mathrm{KA}_i}$, $\widehat{\mathrm{MA}_i}$, and $\widehat{\mathrm{SA}_i}$ is the fitted values of knowledge, market, and supplier access of country i, which are calculated by using coefficients obtained from first stage estimation. We suppose that mobile factors of production v_i are mobile freely, and $v_i = v$ for every i. Therefore, it is included in the constant terms.

Moreover, several econometric issues must be addressed. The first is the endogeneity problem. Variables outside of our model and not included in control variables may correlate with knowledge access and manufacturing wage. In order to avoid this possibility, we estimate the model by a two-step efficient generalized method of moments (GMM), as well as by OLS. For candidates of instrument variables, we use the distance from the three countries United States, Japan, and the European Union[5], and the numbers of television sets, radios, and daily newspapers per 1000 people. In order to test the overidentification condition and econometric model specification, we use Hansen's J statistics. We compute White-robust standard error to correct the heteroskedasticity condition.

Second, high correlations exist between regressors. The correlation coefficient between log market and knowledge access is 0.784, and that between log market and supplier access is 0.941. This implies that we cannot use the market and supplier access together in eq. (14). Our concern is mainly the effect of knowledge spillover. Thus, following Redding and Venables (2004), we remove supplier access from the wage equation, and only use market access as the control variable of manufacturing goods accessibility.

Finally, there is a generated regressor problem, which was pointed out by Pagan (1984). In the two-stage estimation procedure that uses generated regressors, estimated standard errors in OLS and GMM are invalid in general except when the tested hypothesis is that the coefficient is zero. Because our concern is whether the effect of knowledge access on the income variation is statistically different from zero or not, we report the t-values, which are assured validity,

[5]We consider Belgium as the central point of the European Union.

rather than the standard errors. However, the validity of this argument is assured only using single generated regressors in a single regression equation. In our regression, there are some specifications that use two generated regressors in a regression. On such specifications, reported *t*-values may include an upward bias[6].

4.2 The simple specification

In the first place, we estimate the wage equation with the simple specification. The result is shown in Table3.

Table 3: OLS results

ln(GDP per worker)	(1)	(2)	(3)	(4)
Observations	55	55	55	55
Year	95−99	95−99	95−99	95−99
ln(KA)	1.671**			
=ln(DKA+FKA)	(4.81)			
ln(FKA)		1.851**		
		(3.20)		
ln(MA)			0.918**	
=ln(DMA+FMA)			(8.20)	
ln(FMA)				1.068**
				(4.85)
Estimation	OLS	OLS	OLS	OLS
R^2	0.3261	0.1399	0.5061	0.2836
F(·)	23.16	10.23	67.24	23.49
Prob>F	0.0000	0.0023	0.0000	0.0000

Heteroskedasticity-robust *t*-values in parentheses.
*denotes statistical significance at the 10% level. **denotes statistical significance at the 5% level.

Column (1) regresses log GDP per worker on log predicted knowledge access by OLS. The coefficient of knowledge access is positive and statistically significant at the 1% level. Column (2) regresses log GDP per worker on log predicted foreign knowledge access by OLS. The coefficient of that is also positive and statistically significant at the 1% level. Moreover, foreign knowledge access alone explains 10% of the cross country wage variation. These results imply that accessibility to foreign knowledge has a positive impact on the manufacturing wage.

Column (3) regresses log GDP per worker on log predicted market access by OLS. The coefficient of market access is positive and statistically significant at the 5% level. Finally,

[6]Redding and Venables (2004) corrected standard errors by using bootstrapping method. However, we cannot use the method because of missing data of knowledge flows.

column (4) regresses log GDP per worker on log predicted foreign market access by OLS. The coefficient of foreign market access is also positive and statistically significant at the 5% level. These results are consistent with the specifications done by Redding and Venables (2004) .

4.3 Estimation with geographic and political factors

In this subsection, we estimate the wage equation specified with fundamental determinants of the income. We use the same control variables as Redding and Venables (2004), which are divided into two groups. The first is a physical geographic group, which includes hydrocarbons per capita, arable land per capita, fraction land in geographical tropics, and prevalence of malaria. The second is a political group, which includes risk of expropriation or protection of property rights, socialist rule during 1950-1995, and the occurrence of an external war (see Data Appendix). We estimate the wage equation controlling for these factors.

Table 4 presents the result. Column (1) regresses log GDP per worker on log total knowledge access by the two-step GMM controlled by geographical and political factors, and regional dummies. We instrument total knowledge access with the distance from the two main markets and sources of knowledge (the United States and Japan), the numbers of radios, and daily newspapers per 1000 people. These instruments passed the orthogonality condition of Hansen's J statistics. In the first-stage regression, these instruments are highly significant by the F-test in which the null hypothesis is that all coefficients are zero. Thus, we can say that the use of these instruments is justified. The coefficient of total knowledge access is still positive and significant at the 5% level. This result shows total knowledge access still affects manufacturing wage even controlling for physical geography and political factors, and regional dummies.

Next, we decompose total knowledge access into the foreign effect and the domestic effect, and regress log GDP per worker by the two-step efficient GMM and the results are given in column (2). log GDP per worker is regressed by foreign and domestic knowledge access with physical geographic and political factors, and regional dummies. We instrument the foreign and domestic knowledge access with the same instruments as in column (1). The use of these instruments is justified by the same tests as column (1). The coefficient of foreign knowledge

Table 4: Economic geography, physical geography, and political factors

ln(GDP per worker)	(1)	(2)	(3)	(4)	(5)
Observations	47	47	54	51	55
Year	95–99	95–99	95–99	95–99	95–99
ln(KA)	2.571**		1.670**	1.743**	1.400**
=ln(DKA+FKA)	(2.95)		[0.572]	[0.536]	(5.28)
			(2.85)	(3.47)	
ln(FKA)		1.332*			
		(1.72)			
ln(DKA)		0.953**			
		(2.82)			
ln(MA)			0.730**	0.388**	0.140
=ln(DMA+FMA)			[0.560]	[0.282]	
			(4.83)	(2.07)	
ln(SA)					0.078
=ln(DSA+FSA)					
ln(hydrocarbons	0.002	−0.013	–	0.018	–
per capita)	(0.89)	(−0.75)	–	(1.04)	–
ln(arable land	−0.022**	−0.029**	–	−0.017**	–
per capita)	(−3.54)	(−4.45)	–	(−2.70)	–
Fraction land in	−1.562**	−1.037**	–	−1.201**	–
geographical tropics	(−3.77)	(−3.02)	–	(−4.02)	–
Prevalence of	0.310	0.073	–	0.208	–
malaria	(0.42)	(0.11)	–	(0.60)	–
Risk of	−0.238**	−0.159*	–	−0.175**	–
expropriation	(−3.00)	(−1.72)	–	(−2.33)	–
Socialist rule	−0.156	0.011	–	−0.578**	–
1950-85	(−0.44)	(0.02)	–	(−2.41)	–
External war	0.301	0.270	–	0.317	–
1960-85	(0.92)	(0.92)	–	(1.40)	–
Regional dummies	yes	yes	yes	yes	–
Hansen(*p*-value)	0.496	0.185	0.212	0.211	–
Estimation	GMM	GMM	GMM	GMM	OLS
R^2	–	–	–	–	0.3936
F(·)	22.22	19.44	37.58	30.42	27.87
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000

Heteroskedasticity-robust *t*-values in parentheses. Standardized coefficients in square parentheses.
*denotes statistical significance at the 10% level. **denotes statistical significance at the 5% level.

is positively signed and statistically significant with 8.6% of p-value. The result of columns (2) shows that foreign knowledge access positively affects the manufacturing wage.

Note that the regression in columns (2) is misspecifications from our theory because we cannot decompose ln(KA) = ln(DKA) + ln(FKA). In spite of the misspecification, the coefficients of these variables are highly statistically significant. These results imply that foreign knowledge accessibility robustly affects the wage.

In order to analyze the relationship between knowledge and market access, we estimate knowledge and market access simultaneously. Column (3) and (4) presents the result. In column (3), we regress the wage equation, including market access, by two-step efficient GMM controlling with regional dummies. We instrument total knowledge and market access by the distance from United States, EU, and Japan, the numbers of telephone sets, and radios per 1000 people. In column (4), we add physical geographic and political factors as control variables. In both specifications, the knowledge access and the market access are positively signed and statistically significant at the 5% level. In those specifications, they pass the specification test. Because of generated regressors problem, reported t-values possibly has upward bias. However, the reported t-values of coefficients of knowledge and market access are high enough, we thus conclude that knowledge access robustly affects the international income variations by controlling for market access.

In order to measure the relative importance of the knowledge versus market access, we calculate standardized coefficients of knowledge and market access. Standardized coefficients are in square parentheses in column (3) and (4). From column (3), knowledge and market access has a similar standardized effect to manufacturing wage. Moreover, in column (4), standardized coefficients of knowledge access is much larger than market access. These results imply that standardized impact of knowledge access to manufacturing wage is at least equivalent to market access.

Finally, following Redding and Venables (2004), we estimate the wage equation (15) including supplier access by using a theoretical restriction $\psi_1 = 1/\beta$, $\psi_2 = 1/\sigma\beta$, $\psi_3 = \alpha/\beta(\sigma - 1)$.

These conditions imply that

$$\psi_2 = \psi_1/\sigma, \quad \psi_3 = \psi_1\alpha/(\sigma - 1).$$

Selecting values of α and σ, we can estimate the wage equation (15) subject to the restriction. Following Redding and Venables (2004) we choose $\alpha = 0.5$, and $\sigma = 10$, and estimate the wage equation without control variables. The result is shown in column (6). Estimated coefficients of market and supplier access are lower than the result of Redding and Venables (2004).

We estimate the wage equation using another fitted value of knowledge access. Using the fitted value of knowledge access which is derived from the coefficients obtained by the first stage estimation using mail flows as a proxy of knowledge flows (column, 2, in Table 2), we estimate wage equation with the same specifications as Table 4. The results are similar to Table 4. This implies that the positive relationship between knowledge access and manufacturing wage is robust even if we use another fitted value of the knowledge access. For more details, see Appendix A.

5 Concluding remarks

In this paper, we investigated the effect of the accessibility to foreign knowledge on international income distribution using the structural model of NEG. Our results suggest that accessibility to foreign knowledge has a positive influence on the manufacturing wage. For example, foreign knowledge access alone explains 10% of cross-country income variations. Its effect is robust by controlling with fundamental geographic and political factors. Moreover, it is considerably robust by controlling with accessibility to the market.

We use bilateral flight passengers as a proxy for knowledge flows. The result implies that exchange of knowledge by face-to-face communications plays an important role in raising manufacturing wages. Thus, we may say that developing countries should reduce transportation costs from foreign countries as well as trade costs in order to increase per capita income. Even if remote countries reduce transportation costs, however, the distance still greatly impedes the raising of incomes.

Note, however, that our paper is limited in that knowledge generation is exogenous. Knowledge access may affect the generation of new ideas. To explore such an effect is one of the most important extensions of our research.

Appendix A Estimation using mail flows

In the main estimation in subsection 4.2 and 4.3, we used the coefficients from the first stage estimation using the number of flight passengers as a proxy of knowledge flows to obtain the fitted value of knowledge access. This is because we consider that face-to-face communications are the most important knowledge spillover. However, this proxy possibly includes some noises other than knowledge flows such as tourists. Therefore, for the check of robustness of the positive relationship between wage and knowledge access, we derive another fitted value of knowledge access using bilateral mail flows as a proxy of knowledge flows (column, 2 in Table 2). Using this knowledge access, we estimate wage equation with the same specification to the previous subsection.

The result is shown in Table 5. The specifications of each column of Table 5 correspond to Table 4. In each column, coefficients of log knowledge access and log foreign knowledge access close to Table 4, and these are statistically significant at the 5% level. It implies that the positive relationship between knowledge access and manufacturing wage is robust even if we use the amount of mail flows as a proxy of knowledge flows in the first stage estimation. Standardized coefficients in columns (3) and (4), and the results of the estimation with the linear restriction (column, 6) are also similar to Table 4.

Note that in columns (3) and (4), from Hansen's J statistics, the overidentification condition does not seem to be satisfied. However, since there are less than 50 observations, the so called small sample problem may arise. It is known that in small sample estimations, Hansen's J test rejects the orthogonality conditions too often (Hayashi, 2000, Chapter 3). We therefore avoid more detailed discussions about the validity of instruments.

Table 5: Estimation using mail flows

ln(GDP per worker)	(1)	(2)	(3)	(4)	(5)
Observations	46	46	53	50	54
Year	95–99	95–99	95–99	95–99	95–99
ln[KA(MAIL)]	1.921**		1.484**	1.701**	1.296**
	(2.75)		[0.611]	[0.675]	(8.12)
			(2.23)	(3.46)	
ln[FKA(MAIL)]		0.837*			
		(1.88)			
ln[DKA(MAIL)]		0.507**			
		(3.78)			
ln(MA)			0.614**	0.313	0.130
=ln(DMA+FMA)			[0.475]	[0.230]	
			(2.39)	(1.49)	
ln(SA)					0.073
=ln(DSA+FSA)					
ln(hydrocarbons	−0.012	−0.008	–	−0.003	–
per capita)	(−0.56)	(−0.42)	–	(−0.15)	
ln(arable land	−0.023**	−0.026**	–	−0.022**	–
per capita)	(−3.29)	(−6.16)	–	(−3.58)	–
Fraction land in	−1.200**	−0.933**	–	−0.915**	–
geographical tropics	(−3.50)	(−3.11)	–	(−3.34)	–
Prevalence of	−0.595	0.876*	–	−0.288	–
malaria	(−1.04)	(−1.79)	–	(−0.78)	–
Risk of	−0.105	−0.115	–	−0.040	–
expropriation	(−1.01)	(−1.58)	–	(−0.57)	–
Socialist rule	−0.507	−0.258	–	−0.979**	–
1950-85	(−1.37)	(−0.96)	–	(−3.24)	–
External war	0.051	−0.092	–	0.099	–
1960-85	(0.16)	(−0.29)	–	(0.43)	–
Regional dummies	yes	yes	yes	yes	–
Hansen(*p*-value)	0.311	0.302	0.063	0.029	–
Estimation	GMM	GMM	GMM	GMM	OLS
R^2	–	–	–	–	0.4501
F(·)	28.71	68.76	35.90	34.55	65.99
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000

Heteroskedasticity-robust *t*-values in parentheses. Standardized coefficients in square parentheses.
*denotes statistical significance at the 10% level. **denotes statistical significance at the 5% level.

Appendix B Data descriptions

Bilateral trade: The World Bank COMTRADE database.

Bilateral flight passengers: ICAO's On-Flight Origin and Destination.

GDP per worker: The current prices (US dollars) GDP and the number of workers is from World Development Indicators 2004.

Bilateral distance: Western Hemispheric Research Resources:
http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/

Risk of expropriation: extent of protection of property rights, measured on a scale from 1 to 5, where a higher score indicates weaker protection of property rights:
http://www.cid.harvard.edu/ciddata.ciddata.html.

Physical geography and institutional, social, and political characteristics: The hydrocarbons (deposits of petroleum and natural gas) per capita, fraction of land area in the geographical tropics, prevalence of malaria, socialist rule, and the occurrence of an external war are from Gallup et al. (1999):
The data can be downloaded from http://www.cid.harvard.edu/ciddata.ciddata.html.

Radios, daily newspapers, and television sets per 1000 people: These data are from World Development Indicators 2004.

South Asia: Bangladesh, India, Nepal, Pakistan, and Sri Lanka.

Southeast Asia: China, Hong Kong, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Thailand, and Taiwan.

Latin America and the Caribbean: Argentina, Bolivia, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, Mexico, Panama, Peru, Uruguay, and Venezuela.

Other Countries: Australia, Austria, Bulgaria, Canada, Denmark, Egypt, Spain, Finland, France, Germany, Greece, Ireland, Iran, Italy, Jordan, Kuwait, Netherlands, Norway, New

Zealand, Poland, Portugal, Romania, Sweden, Turkey, United Kingdom, United States, and South Africa.

References

[1] Baldwin, R.E., Ph. Martin (2004) Agglomeration and Regional Growth, J.V. Henderson, J.-F. Thisse, ed., *Handbook of Regional and Urban Economics*, 4, 2671-2711, Elsevier, Amsterdam.

[2] Baldwin, R.E., Ph. Martin, G.I.P. Ottaviano (2001) Global Income Divergence, Trade, and Industrialization: The Geography of Growth Take-Offs, *Journal of Economic Growth* 6: 5-37

[3] Coe, D., E. Helpman (1995) International R&D Spillovers, *European Economic Review* 39, 859-887

[4] *Data Book of The World 2004* (2004) Ninomiya Shoten, Tokyo (in Japanese)

[5] Fujita, M., J.-F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge.

[6] Grossman, G., E. Helpman (1991) *Innovation and Growth in the World Economy*, MIT Press, Cambridge.

[7] Hanson, G. (2005) Market Potential, Increasing Returns and Geographic Concentration, *Journal of International Economics* 67, 1: 1-24.

[8] Hayashi, F. (2000) *Econometrics*, Princeton University Press, Princeton.

[9] Head, K., T. Mayer. (2004a) The Empirics of Agglomeration and Trade, J.V. Henderson, J.-F. Thisse, ed., *Handbook of Regional and Urban Economics*, 4, 2609-2670, Elsevier, Amsterdam.

[10] ———— (2004b) Market Potential and the Location of Japanese Investment in the European Union, *The Review of Economics and Statistics* 86, 4: 959-972.

[11] Keller, W. (2002) Geographic Localization of International Technology Diffusion, *American Economic Review* 92, 1: 120-142

[12] Krugman, P. (1980) Scale Economies, Product Differentiation, and the Pattern of Trade, *American Economic Review* 70: 950-959.

[13] ——— (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99, 3: 483-499.

[14] Pagan, A. (1984) Econometric Issues in the Analysis of Regressions with Generated Regressors, *International Economic Review* 25, 1: 221-247

[15] Peri, G. (2005) Determinants of Knowledge Flows and Their Effect on Innovation, *The Review of Economics and Statistics* 87, 2: 308-322

[16] Redding, S., A.J. Venables. (2004) Economic Geography and International Inequality, *Journal of International Economics* 62, 1: 53-82.

[17] World Bank (1995) *World Development Report 1995*, Oxford University Press, New York.