경제학 석사학위 논문

Education, Distance to Frontier and Economic Growth

- A Critical Review of Empirical Analyses -

교육, 기술발전과 경제성장

-실증 분석을 중심으로-

2019년 7월

서울대학교 대학원
경제학부 경제학 전공
서 바 로

Education, Distance to Frontier and Technological Growth

- A Critical Review of Empirical Analyses -

지도교수 김 세 직

이 논문을 경제학 석사학위 논문으로 제출함

2019년 4월

서울대학교 대학원
경제학부 경제학 전공
서 바 로

서바로의 석사학위 논문을 인준함
2019년 7월

위 원 장	김 영 식	(인)
부 위 원 장	김 세 직	(인)
위 원	김 소 영	(인)

Abstract

We discuss a potential problem inherent in an empirical analysis of higher education and subsequent technological growth. By following the results of existing literature, we assume that tertiary education would play a more prominent role in advanced economies, and we empirically investigate this hypothesis empirically, specifically by using an interaction term on tertiary education attainment and proximity to the technological frontier. We however find that the results diverge from existing literature if we use different timelines and different samples in the estimation. In particular, by including dummy variables in the regression, we find that the proclaimed positive relation between tertiary education and technology growth does not hold in the periods in which countries experience rapid economic growth. Consequently, we claim that a reconsideration of the model is necessary if the validity is to be universally maintained.

Keyword : Tertiary Education, Technology Growth, Composition of Human Capital, Distance to Frontier
Student Number : 2016-29711

Table of Contents

Chapter 1. Introduction

The purpose of this thesis is to discuss a potential problem with an empirical analysis of education and technological growth. Specifically, we deal with the inconsistency shown in results of the regression analysis made by Vandenbussche, Aghion and Meghir(2006) (henceforth VAM). Their analysis has been known to provide convincing evidence supporting the theory that technology growth in a country can be induced by different tiers of education, depending on how advanced their economy would be. However, we find that their method provided erratic outcomes depending on the sample and timeline used in the analysis. Using a specification method, made possible by employing an updated dataset with a larger sample, we observe how significance and magnitude of the coefficients change if we make adjustments to their initial sample. All in all, we propose that the consistency of the VAM(2006) model should be reconsidered if its validity is to be universally maintained.

VAM(2006) has been regarded pivotal because it was one of the first studies to provide significant empirical evidence supporting the theory of endogenous growth. Their analysis was based on the idea that higher education, and the subsequent formulation of high skilled labor, should be more important in countries with more advanced economies. According to the idea, as countries approach the technological frontier, there would no longer be technology from other nations to imitate and adopt. Rather, the countries would need to enhance their own technology with innovation, which would be attainable only with high-skilled human capital. Through an empirical analysis with panel data consisting of 19 OECD countries from 1960-2000, VAM(2006) used an interaction term between 'the fraction of the population with tertiary education' and 'proximity to the technological frontier' to

capture this effect. The results of their analysis were deemed statistically significant and consistent with their initial hypothesis. Thus, subsequent studies such as Aghion et al.(2009) (henceforth AHV), Ha, Kim and Lee(2009) (henceforth HKL) and Islam(2010) have built upon their framework and applied it on diverse datasets. Also, as stated in VAM(2006), it had even been considered possible for policy makers to endorse and apply the results to determine the optimal composition of education expenditure in their countries.

In this thesis, however, we suggest that a precaution should be necessary before their results could be globally accepted. Specifically, we point out that the dataset they employed may have some inherent problems, especially, as acknowledged in ABHV(2009), with the fact that theirs was dubiously small. Sampling a mere 19 countries among 30 OECD members,[①] although it might have been inevitable due to insufficient available data, seems to have allowed the outcomes to avoid severe consequences that would have resulted otherwise. Most of all, we observe that adding new countries into the dataset seriously distort the estimation results. Also, we find that an addition, or reduction, of certain time periods included in the analysis generate conflicting outcomes. Furthermore, we are not only able discover such insufficiencies, but also to isolate observations that appear to be causing the inconsistencies.

The thesis is organized as follows: Chapter 2 briefly reviews the literature preceding, as well as following VAM(2009). Chapter 3 describes the data used in the analysis. Chapter 4 introduces the specification methods and provides the empirical results. Chapter 5 summarizes with a brief conclusion.

---

① Their analysis was based on the data of the countries available up until 2000.
In 2019, however, there are 36 members in the OECD: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the UK and the USA.
The underlined are those countries that joined after 2000.

Chapter 2. Related Literature

VAM(2006) presented a two-part paper with both a theoretical and an empirical part. In the theoretical section, they established a model in which labor is stratified into low-skilled and high-skilled labor. It was assumed that low-skilled labor is associated with imitating and adopting existing technology while high-skilled labor promotes innovation. They claimed that technological growth is induced by some combination of both kinds of labor, but high-skilled labor becomes more dominant in the combination as a country moves closer to the technological frontier.

The theoretical model presented in VAM(2006) was actually built upon a series of models from existing literature. Such literature include that of Nelson and Phelps(1966), who first formally presented the idea that education stimulates technology adoption. Alongside this 'technology-adopting role', the 'innovative role' of education was also emphasized. Acemoglu, Aghion and Zilibotti(2006) (henceforth AAZ) was such a study in which innovation, alongside imitation, plays a decisive role for technological progress. VAM(2006) was built directly on this model, complementing it by introducing low and high-skilled labor to respectively be the driving forces behind each engine of growth, imitation and innovation.

Unlike theory, empirical studies preceding VAM(2006) were relatively less successful in presenting promising results. One possible explanation of such insufficiency is the measurement error apparent in existing educational data. Kreuger and Lindahl(2001) presented evidence that, when using data subject to measurement error, regression on education and growth could produce attenuated estimates, possibly explaining why empirical literature such as Benhahib and Spiegel(1994) had been providing some insignificant results on

the subject. Kreuger and Lindahl(2001) also conducted their own analysis after carefully accounting for measurement error. Their discovery was that education affects growth significantly only in countries with the lowest level of initial education.

The empirical section in VAM(2006) presented a possible solution to the puzzle of Kreuger and Lindahl(2001). Their significant results and also the solid theory underlying the hypothesis have led subsequent literature to endorse their framework and apply it to a variety of datasets. Such proponents include ABHV(2010), HKL(2009) and Islam(2010). ABHV(2010), in pursuit of revealing a causal effect, have applied the same empirical framework to U.S. states, ergo enriched with a larger sample and more possible IV. HKL(2009) also used the same method, with measures for R&D investment in place of education attainment data, on 3 east Asian countries: Japan, Korea and Taiwan. Islam(2010) expanded the sample to include 87 countries, grouped into 3 income groups.

This study differs from such preceding studies because we start our analysis with a case in which the proposed assertion does not hold. We consider all possibilities and propose an explanation to why such an inconsistency exists.

Chapter 3. Data

3.1. Data Source

We utilize two data sets in this analysis. First, we use Penn World Table 9.1 (henceforth PWT)[②], a cross-country data set which provides abundant yearly cross-country observations for 182 countries from 1950 to 2017. Second, schooling data were collected from the Barro-Lee 2.0 datase[③](henceforth BL) which provides educational attainment estimates for 146 countries from 1950 to 2010 in 5-year intervals.

All data sources used are the same to, but updated versions of, the ones that were used in VAM(2006). Data sets we were not able to utilize are (1) the educational data set from De la Feunte and Domenech(2002) (henceforth DD), which was used as an alternative source for a measure of educational attainment, and (2) the UNESCO yearbooks which were used to extract public expenditure data.

The reason we could not utilize the DD dataset is because it is no longer available.[④] Therefore, in our analysis, we only compare results that made use of educational data from BL2.0. Meanwhile, we decided not to extract public expenditure data from the UNESCO yearbooks because doing so would reduce the size of attainable sample countries to a mere 19. The UNESCO data were used as instruments in the original analysis, but unfortunately, there are no other sources that provide the same data dating back to 1950. Therefore, we settled to use tertiary education data lagged two

② Last updated on April 30, 2019
③ Last updated on June 2018
④ https://ideas.repec.org/p/fda/fdaddt/2014-14.html

periods as an alternative instrumental variable for tertiary education.

We construct our panel with countries that provide sufficient data in both our data sources. In the original analysis, a total of 19 OECD countries were used.⑤ The reason OECD countries were used is because the authors wished to establish a valid pattern among developed countries. In this sense we also restrict our analysis to OECD countries. Our final panel consists of 28 OECD countries⑥ with a timeline from 1955-2010.

The following tables provide summary statistics for the countries in our sample.

Table I. Income and Average Growth

	(1) Current Income (2017, US$, per capita)		**(2) Initial Income** (1950, US$, per capita)		**(3) Average Growth** (1950 - 2017)	
Average		44020.75		9156.05		2.50%
1	NOR	83497.27	CHE	21442.29	KOR	5.10%
2	IRL	74167.63	NOR	16337.87	JPN	3.73%
3	CHE	62178.28	USA	14436.85	IRE	3.52%
4	LUX	56228.93	AUS	14250.03	ESP	3.03%
5	USA	54586.24	NZL	13178.38	DEU	2.94%
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	ISR	32604.44	GRC	4456.60	AUS	1.86%
25	TUR	25583.35	ESP	4384.83	CAN	1.85%
26	PRT	24582.38	PRT	3512.52	MEX	1.84%
27	GRC	21796.29	JPN	3215.43	CHE	1.58%
28	MEX	16792.37	KOR	1407.33	NZL	1.50%

⑤ Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Ireland, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Switzerland, UK and USA

⑥ Added Countries : Chile Germany, Iceland, Israel, Japan, Korea, Luxembourg and Mexico and Turkey. The OECD countries that were not included in the analysis, due to insufficient data are : Czech Republic, Estonia, Hungary, Latvia, Lithuania, Slovenia,

[Table I], compares the 2017 and 1950 incomes of the countries in our panel. Column [I-(1)] shows that all 28 countries are highly wealthy countries. In fact all countries have an income of over $20,000 per capita if we exclude Mexico. However, as seen in column [I-(2)], not all countries were wealthy from the beginning. Take for example Korea and Japan. In 1950, Korea and Japan were the poorest countries among the 28. However, as seen in column [I-(3)] they maintained the highest average growth rates among the 28. In 2018, Korea has around $37,000 per capita income while Japan is experiencing around $39,000 per capita income. This shows that even among the wealthy OECD countries there is considerable variation in growth patterns among the countries.

Table II. Productivity Growth and Proximity to the Frontier

	(1) TFP Growth (average, 1950 - 2017)		**(2) TFP Growth2** (recent, 2012 - 2017)		**(3) Proximity** (2017)	
Average		0.86%		0.44%		0.85
1	IRL	1.67%	IRE	3.92%	IRL	1.10
2	DEU	1.56%	ISL	1.70%	NOR	1.01
3	KOR	1.54%	CAN	1.08%	USA	1
4	FRA	1.42%	DEU	1.01%	DEU	0.97
5	ISR	1.32%	AUS	0.98%	FRA	0.96
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	LUX	0.39%	LUX	-0.08%	JPN	0.65
25	NZL	0.32%	NZL	-0.12%	MEX	0.64
26	TUR	0.13%	GRC	-0.20%	KOR	0.63
27	CHL	0.07%	CHL	-0.51%	PRT	0.60
28	MEX	-0.46%	PRT	-1.02%	GRC	0.53

[Table II] provides summarized statistics on measures for the productivity growth rates and proximity. PWT9.1 provides rTFPna, and cTFP data on, as measures of productivity. Both measures use equation (1) to compute the

relative productivity $(\boldsymbol{A_j A_k})^{\frac{1}{2}}$ of economy $\boldsymbol{j}$ compared to that of $\boldsymbol{k}$ ⑦

$$(A_j A_k)^{\frac{1}{2}} = \left(\frac{GDP_j}{GDP_k}\right) / Q_T \tag{1}$$

Specifically, rTFPna uses the time series of rGDPna for each country and uses the GDP in 2011 as the base(GDP_k). Evidently, rTFPna measures the productivity each year compared to that of 2011. Meanwhile, cTFP uses the cGDP data (measured in PPP) to compute the productivity. Here, the cGDP of the U.S is used as the base. Analogous to rTFPna, cTFP measures the productivity of a country, relative to the U.S. Therefore, rTFPna can be used as a variable that measures productivity growth. cTFP can be used to measure the proximity to the world technological frontier.

In column [II-(1)] and [II-(2)] the average productivity growth rates, in 1950-2017 and 2012-2017 respectively, are given. It can be seen that there is little variation in the order, but the overall growth rates are decreasing in each country.

Meanwhile in column [II-(3)] the proximity to the technological frontier for each country is given. Here it is noticeable that the countries with more proximate levels tend to grow faster, as seen in columns [II-(1)] and [II-(2)]. Also we can see that the countries with the lowest level of proximity tend to be in similar geological regions *i.e.* Asia and Southern Europe.

⑦ The computational method is described in detail in Feenstra et al.(2013).

3.2. Variable Description

The following are brief descriptions of the variables used in the analysis. To account for business cycle fluctuations and focus on long term growth, we take 5 year moving averages on all yearly data provided by PWT9.1. Due to availability of the BL2.0 data we construct our panel in 5-year intervals, starting from 1955.

(1) Total Factor Productivity(TFP)

We use the Solow residual as Total Factor Productivity(TFP).
First we assume that, in any arbitrary country, aggregate production function follows the Cobb-Douglas function.

$$Y_{j,t} = z_{j,t} K_{j,t}^{1-\alpha} L_{j,t}^{\alpha} \tag{2}$$

K stands for aggregate capital stock, L stands for aggregate labor, Y is the gross domestic product and α is the labor share of income. Note that z can be interpreted as productivity of country j at time period t because it shows how much gross product can be produced with inputs of K and L.

TFP, and TFP growth can be computed as follows.

$$\begin{aligned} \ln(Y_{j,t}) &= \ln(z_{j,t}) + (1-\alpha)\ln(K_{j,t}) + \alpha \ln(L_{j,t}) \\ \ln(z_{j,t}) &= \ln(Y_{j,t}) - (1-\alpha)\ln(K_{j,t}) + \alpha \ln(L_{j,t}) \\ \Delta z_{j,t} &= \ln(z_{j,t}) - \ln(z_{j,t-1}) \end{aligned} \tag{3}$$

Data corresponding to Y, K, L are all given in PWT 9.1. Although PWT9.1 also provides values for labor share of income α, we follow the method of VAM(2006) and apply 0.7 to all values of α.

(2) Proximity to the Technological Frontier

Next we define a variable that measures how proximate a country is to the technological frontier. The underlying assumption is that the United States of America has the leading technology in the world. This assumption is reasonable because, (1) Many U.S. based companies are often regarded as the technology leader in diverse industries and (2) because the U.S. is also the main source of technological diffusion, being trade partners with most countries in the sample.

In this sense, our measure of proximity for arbitrary country j is the following.

$$Proximity \equiv \ln\left(\frac{A_j}{A_{US}}\right)$$

In place of A_j, the productivity level of country j, we may use the TFP computed in the last section. Note that normally, since the U.S. is defined to be the world technological frontier, $A_j < A_{US}$ *i.e.* the measure would be a negative value.

(3) Educational Attainment

We use the BL2.0 dataset to accumulate cross-country data on fractions of the population that have received tertiary, secondary and primary education. We define 'high' education as the population that have received at least some tertiary education.

The population of reference we use is adults between 15 and 64. We use 15+ instead of 25+ in order to include the population between 15 and 25 who joined the labor force at an early age without college education.

[Table III] provides descriptive statistics for the main variables of interest.

Table III. Descriptive Statistics for Main Variables

	mean	std.dev	min	25%	50%	75%	max
PWT 9.1							
ΔTFP	0.016	0.02	-0.08	0.005	0.017	0.029	0.073
Proximity	-0.304	0.307	-1.39	-0.45	-0.29	-0.12	0.48
BL 2.0							
f^h	14.06%	11.13%	0.95%	5.03%	11.10%	20.61%	57.28%
f^l	77.1%	14.03%	22.62%	69.76%	78.95%	88.03%	97.84%

Chapter 4. Empirical Analysis

4.1. Baseline Model

The following equation is the baseline regression model used in our empirical analysis.

$$\Delta TFP_{j,t} = \beta_1 A_{j,t-1} + \beta_2 f^h_{j,t-1} + \beta_3 \left(A_{j,t-1} \times f^h_{j,t-1}\right) + v_t + u_j + \varepsilon_{j,t} \qquad (5)$$

$\Delta TFP_{j,t}$ stands for TFP growth, $A_{j,t-1}$ stands for proximity and $f_{j,t-1}$ stands for the fraction of population that received higher education in country j. Note that we use lagged terms for f and A, considering the existent time lag between an individual receiving education and entering the labor market. v_t stands for the time variant conditions that affect the economy worldwide. u_j stands for time invariant country specific effects.

Using the regression model (5), we may test the following hypotheses.

[Hypothesis1]. As a country moves closer to the technological frontier, high-skilled labor becomes more important for subsequent growth.

[Hypothesis2]. (Technological Convergence) Technology grows at a slower pace in countries that are closer to the technological frontier.

[Hypothesis3]. Countries with an abundant level of high-skilled workers experience less reduction in growth rates as they approach the technological frontier.

[Hypothesis1] can be verified by checking β_3, the coefficient for the interaction term between 'proximity' and 'tertiary education attainment'. [Hypothesis2] holds if β_1, the coefficient for proximity, is significant and negative. β_1 being negative technically means that $\Delta A_{j,t-1} > 0$ would lead to $\Delta(\Delta TFP_{j,t})$ being negative, *i.e.* the growth rate of TFP slows down as the proximity level increases. Finally, [Hypothesis 3] is true if both [Hypothesis1] and [Hypothesis2] hold simultaneously.

4.2. Specification Methods

In order to discuss the specification model in detail, we start by performing a baseline regression and present the results. The base line regression is performed on a panel of 19 countries⑧, 1960-2010

[Table IV] reports the estimates of 5 regressions conducted on the same panel with 19 countries from 1960-2010. In all 5 regressions, time dummies were included in order to control for v_t observed in equation (5). Also, all reported standard errors allow for serial correlation and heteroskedasticity, by being estimated with clusters by country.

Differences among the 5 regression results stem from whether or not we included the interaction term, and whether or not we controlled for country specific fixed effects.

⑧ The original 19 countries used in VAM(2006)

Table IV. Regression on 19 Countries, 1960-2010

VARIABLES	(1) ΔTFP	(2) ΔTFP	(3) ΔTFP	(4) ΔTFP	(5) ΔTFP
A	-0.012**	-0.080***	-0.018**	-0.089***	-0.030***
	(0.005)	(0.016)	(0.007)	(0.020)	(0.010)
f^h	0.017*	0.036	0.023**	0.040	0.034*
	(0.010)	(0.053)	(0.010)	(0.045)	(0.018)
$A \times f^h$			0.045	0.073	0.095**
			(0.039)	(0.073)	(0.047)
Observations	171	171	171	171	171
Number of Countries	19	19	19	19	19
Year FE	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	Group

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In both our analysis and VAM(2006), instrumental variables on within groups is used as the specification method. In our analysis $A_{j,t-2}$ and $f^h_{j,t-2}$ are used as IV for $A_{j,t-1}$ and $f^h_{j,t-1}$. The analysis in VAM(2006) differed slightly by using 'government expenditure in tertiary education'⑨ as an IV for $f^h_{j,t-1}$.

According to ABHV(2009), both IVs used in each regression are actually insufficient to establish a causal relation. The insufficiency is due to some omitted variables being correlated with either IV. In an attempt to reduce this correlation we include time and country fixed effects in the model.

As stated above, we include fixed effects in this model to overcome some of the endogeneity bias that stems from omitted variables. But, as we can observe from comparing the results in columns], the significance of the

⑨ Lagged 2 terms.

estimation results is greatly reduced if we include country fixed effects in the regression. We suspect that the circumstance is probably due to the inherently insufficient spending variation in the data we work with. Therefore, we group the countries into 11 groups according to geographical and/or industrial proximity[10], and use grouped fixed affects in the regression. By doing so, we are able to simultaneously control for the fixed effects that may distort the results while also maintaining some variation.

4.3. Empirical Results

(1) Regression on Diverse Samples of Countries

[Table V] compares the estimates from regressions on diverse samples. We present only the results from regressions allowing grouped fixed effects due to reasons stated in [Section 4.2]. The results of all regressions are presented in the appendix.

In column [V-(2)] we add all 28 OECD countries from our sample into the regression. By doing so, we observe that the estimation results diverges sharply from the initial one, [V-(1)]. By arbitrarily removing and adding countries from the sample and performing multiple regression, we find that some countries are especially responsible for the distortion in results. We group countries according to the effects they bring when included into the sample.

The first group of countries, Chile, Greece and Korea, are countries that singularly changes the whole outcome of an estimation. The second group of countries, Japan and Turkey, do not bring glaring bias into the estimation, but

[10] The groups are provided in the appendix.

do attenuate the results. Finally, the third group of countries, Germany, Iceland, Luxembourg and Mexico are countries, opposed to group1 and group2, that instead reinforce the results when included in the panel.

Table V. Regression on Diverse Samples of Countries (1960-2010)

VARIABLES	(1) ΔTFP	(2) ΔTFP	(3) ΔTFP	(4) ΔTFP
A	-0.030***	-0.024***	-0.032***	-0.035***
	(0.010)	(0.004)	(0.009)	(0.010)
f^h	0.034*	0.031*	0.028	0.040**
	(0.018)	(0.018)	(0.018)	(0.017)
$A \times f^h$	0.095**	-0.028	0.076*	0.101**
	(0.047)	(0.051)	(0.045)	(0.045)
Observations	171	252	225	207
Number of Countries	19	28	25	23
Year FE	Yes	Yes	Yes	Yes
Country FE	Group	Group	Group	Group

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The result of using these diverse panels in the regression is depicted well in [Table V]. Column [V-(2)] shows how the estimates change if we add all 9 new countries into the estimation. We can observe that β_4, the coefficient for the interaction term becomes both insignificant and also negative in this regression. So we remove countries from group 1 and perform a new regression, as in column [V-(3)]. Here the reported magnitude and significance of the coefficients return to being consistent with the hypotheses. But we recognize there is still some room for improvement compared to the results of column [V-(1)]. Finally, by additionally removing members of group2, The results become more significant compared to the initial ones, as can be seen in column [V-(4)].

We now interpret the results in [Table V]. First, we notice that regardless of the sample used, [Hypothesis 2] holds in all cases. In other words, all countries in the sample tend to experience slower technological progress as they come closer to the frontier. Considering that all 28 countries are members of the OECD with highly advanced economies, the possible convergence to TFP frontier is a reasonable observation. The second point to notice is that countries in group 1 and group 2 seem to diverge from the existing trend. While 23 of the 28 countries show consistence to all 3 hypotheses, the remaining 5 countries, especially the ones in group1, certainly do not follow the patterns of [Hypothesis 1] and [Hypothesis3].

(2) Regression with Diverse Timelines

Now we fix the sample size to the 23 countries that have shown consistence for all 3 hypotheses, and perform the regressions with diverse timelines. We intend to observe the fluctuations in midst of altering the time periods.

We start by excluding the earlier stages in the timeline. In [Table VI], we can observe the changes in outcomes when we gradually exclude the earlier time periods. We can observe that the absolute value and significance of coefficients tend to decrease as further stages are excluded

Meanwhile in [Table IV], we remove the latest time periods from the estimation model. Here we may observe that the results are irregular depending on the time period we use. We cannot identify any clear patterns.

Table VI. Regression with Diverse Timelines (23 Countries, -2010)

	1955-2010	1960-2010	1965-2010	1970-2010	1975-2010
	(1)	(2)	(3)	(4)	(5)
VARIABLES	ΔTFP	ΔTFP	ΔTFP	ΔTFP	ΔTFP
A	-0.032***	-0.035***	-0.019	-0.023*	-0.020
	(0.010)	(0.010)	(0.014)	(0.013)	(0.014)
f^h	0.038**	0.040**	0.014	0.024	0.010
	(0.017)	(0.017)	(0.017)	(0.016)	(0.018)
$A \times f^h$	0.102*	0.101**	0.022	0.040	0.005
	(0.053)	(0.045)	(0.055)	(0.051)	(0.051)
Observations	230	207	184	161	138
Number of Countries	23	23	23	23	23
Year FE	Yes	Yes	Yes	Yes	Yes
Country FE	Group	Group	Group	Group	Group

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table VII. Regression with Diverse Timelines (23 Countries, 1955-)

	1955-2010	1955-2005	1955-2000	1955-1995	1955-1990
	(1)	(2)	(3)	(4)	(5)
VARIABLES	ΔTFP	ΔTFP	ΔTFP	ΔTFP	ΔTFP
A	-0.032***	-0.030***	-0.032**	-0.039***	-0.037**
	(0.010)	(0.011)	(0.014)	(0.014)	(0.016)
f^h	0.038**	0.047**	0.042*	0.071***	0.015
	(0.017)	(0.018)	(0.023)	(0.026)	(0.040)
$A \times f^h$	0.102*	0.085	0.120	0.228**	0.250
	(0.053)	(0.060)	(0.103)	(0.114)	(0.161)
Observations	230	207	184	161	138
Number of Countries	23	23	23	23	23
Year FE	Yes	Yes	Yes	Yes	Yes
Country FE	Group	Group	Group	Group	Group

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table VIII. Regression with Diverse Timelines (23 Countries, 1960-)

	1960-2010	1960-2005	1960-2000	1960-1995	1960-1990
	(1)	(2)	(3)	(4)	(5)
VARIABLES	ΔTFP	ΔTFP	ΔTFP	ΔTFP	ΔTFP
A	-0.035***	-0.033***	-0.036**	-0.048***	-0.048***
	(0.010)	(0.011)	(0.015)	(0.014)	(0.018)
f^h	0.040**	0.053***	0.051**	0.091***	0.034
	(0.017)	(0.018)	(0.022)	(0.026)	(0.039)
$A \times f^h$	0.101**	0.081	0.124	0.281**	0.338**
	(0.045)	(0.057)	(0.105)	(0.110)	(0.171)
Observations	207	184	161	138	115
Number of Countries	23	23	23	23	23
Year FE	Yes	Yes	Yes	Yes	Yes
Country FE	Group	Group	Group	Group	Group

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We try excluding more time periods from the estimation, while being cautious not to lose too many observations. For example, [Table V] is what happens when we use 1960 as the starting time period instead of 1955. The results are similar to that of [Table IV], but it can be recognized that the coefficients β_1 and β_3 are much more significant with a larger magnitude.

4.3. Further Specification : Including Dummy Variables

In this section, we include dummy variables into the regression in order to identify the pattern underlying the phenomenon presented in the previous section. We start of by grouping the observations into 4 groups according to the GDP per capita and long-term GDP growth the country has shown in that particular period.

Table IX. Descriptive Statistics for Growth Patterns

	mean	std.dev	min	25%	50%	75%	max
28 Countries (336 observations, 1955~2010)							
GDP($) *per Capita*	23659.8	13399.97	1558.32	13777.70	21641.63	31763.77	81692.91
ΔGDP *5year MA*	0.03642	0.02488	-0.0619	0.0223	0.0348	0.0503	0.11087

[Table IX] shows the descriptive statistics of growth patterns shown in the 28 countries included in the panel. We may observe that throughout the timeline, most of the countries have averaged a steady long-term growth path of 2~3% and have reached a per capita income above $30,000 in the 2000s.

However, some countries show divergent patterns. As also seen in [Table II], some countries that have high per-capita income today, have reached such status only after following a rapid growth path of 5~7% long-term growth for almost 20 years. By using dummy variables, we wish to separate the heterogenous effects such diverse growth patterns have on technology growth.

First, we define countries to have 'high income' if a country has over $30,000[11] GDP per capita in a certain period. Similarly we define 5%[12] long-term growth as a signal for 'rapid growth'. By interacting these two definitions we can divide each observation into the following 4 groups:

Group1. High Income, Rapid Growth. 9 observations
Group2. High Income, Steady Growth 89 observations
Group3. Low Income, Rapid Growth 78 observations
Group4. Low Income, Steady Growth 160 observations

[11] The Q3 observation of GDP per capita
[12] The Q3 observation of GDP growth (5 year moving average)

[Group 1] and [Group 2] respectively group the countries according to the growth patterns they show after a certain level of per capita income has been reached. If an observation is categorized as [Group 1], it means that the country is simultaneously experiencing both high living standards and rapid long-term growth. This is in fact an extremely rare case and mainly consists of Scandinavian countries in the 2000s. Meanwhile, being in [Group 2] means that the country is retaining a steady, or slow, growth path after it has reached a certain GDP level. The group consists of most developed countries after the 1990s.

[Group 3] and [Group 4] divide the countries according to the growth patterns they have shown in their developing stage. For example, [Group 3] consists of countries such as Korea and Japan in the 1970s and 1980s. These countries were initially endowed with a very small GDP level in the 1950s but have experienced rapid growth in subsequent decades. However, these instances are actually the exception, while most OECD countries have steadily grown on average 2% annually. These instances are grouped in [Group 4].

By assigning dummy variables to each group and performing regression as in equation (5), we are able to observe how the patterns of education and technology growth is heterogenous according to the growth pattern the country is following. [Table X] shows the regression results. We can observe in column (1) that the coefficient for $A \times f^h$ is indeed not positive nor significant when we include all 28 countries in the regression. However, as seen in column (2), we may observe from the coefficient of $A \times f^h \times Group4$ that if we exclude exceptional growth patterns, the effect becomes both positive and significant. This shows that the tertiary education becomes increasingly important for countries with advanced technology, only if it is steadily developing under a growth pattern depicted by traditional growth theory.

Table IX. Regression with Dummy Variables

(23 Countries, 1955~2010)

VARIABLES	(1) ΔTFP	(2) ΔTFP
A	-0.021***	-0.012*
	(0.005)	(0.007)
f^h	0.027	0.017
	(0.020)	(0.019)
$A \times f^h$	-0.033	-0.454***
	(0.055)	(0.161)
$A \times f^h \times Group2$		0.015
		(0.055)
$A \times f^h \times Group3$		0.167
		(0.164)
$A \times f^h \times Group4$		0.437***
		(0.150)
Observations	280	280
Number of Countries	28	28
Year FE	Yes	Yes
Country FE	Group	Group

There are a few more facts that may be inferred from [Table X]. First of all, it can be seen that when countries reach a certain income level, the coefficient is insignificant, or even negative and significant. Seeing that most countries reach a per capita income level higher than $30,000 in the 2000s, it can be inferred that recent technology development can be less reliant on tertiary education. Also, the coefficient of $A \times f^h \times Group3$ being insignificant may be due to the fact that tertiary education operates differently in countries that are growing rapidly. All in all, we may observe that the pattern described in VAM(2006) may not be maintained globally.

Chapter 5. Concluding Remarks

In this thesis, we have used panel data analysis with interaction terms to specify the effects higher education has on technological growth, in midst of growing proximity to the technological frontier. The specification method we have used is instrumental variables on within groups, in order to possibly mitigate the potential endogeneity in the regression.

Our findings can be summarized as follows. The result of our estimation is significant and consistent to our hypotheses when we perform regression on only a small group of OECD countries. The consistency to our hypotheses imply that indeed the more proximate the country is to the technological frontier, the more prominent the role of education becomes for subsequent TFP growth. However, we also find that when we augment the sample with more high income OECD countries, the results become increasingly insignificant. In addition, if we use different timelines for the panel regression, there are erratic outcomes.

There are three possible explanations for this phenomenon. First, this can be due measurement error of tertiary education attainment data, and the consequential endogeneity. Endogeneity might have attenuated the estimates, or may even have caused a spurious positive association between the variables. Second, the inconsistency may be due to small subgroups of time periods and countries that go against the trend shown in other OECD countries. In fact we have been able to identify the time periods and the countries responsible. But we have not been able to determine a conspicuous pattern, and thus have not been able to provide a coherent explanation for our results.

A pressing extension to our thesis is to identify the exact cause for the

phenomenon. For example we could accumulate data from the subgroup of countries that have found to be distorting the estimation results. By doing so we may be able to find similarities between the subgroup and proclaim a clear pattern. Then, we would be able to formulate a model that can be maintained gloabally.

References

[1] Acemoglu, D., Aghion, P., and Zilibotti, F. (2006), "Distance to frontier, Selection and Economic Growth", *Journal of the European Economic Association*, 4(1):37-74.

[2] Aghion, P., Boustan, L., Hoxby, C., and Vandenbussche, J. (2009), "The Causal Impact of Education on Economic Growth: Evidence from US", *Brooking Papers on Economic Activity*, 1:1-73.

[3] Aghion, P. and Howitt, P. (2006), "Joseph Schumpeter Lecture, Appropriate Growth Policy: A Unifying Framework", Journal of the European Economic Association, 4(2-3):269-314.

[4] Barro, Robert and Lee J-W. (2000), "International Data on Educational Attainment: Updates and Implications", *Working Paper no.42, Center for International Development at Harvard University*, April 2000.

[5] Barro, Robert J. (1998) "Human Capital and Growth in Cross-Country Regressions", *mimeo, Harvard University*, October 1998.

[6] Benhabib, J. and Spiegel, M. (1994), "The Role of Human Capital in Economic Development: Evidence from Aggregate Cross Country Data", *Journal of Monetary Economics*, 34(2): 143-174.

[7] Caselli, F. and Coleman, W.J. (2001), "Cross-country technology diffusion: The. Case of Computers", *American Economic Review*, 91(2):328-335.

[8] Ciccone, A., Papaioannou, E. (2009), Human Capital, the Structure of

Production and Growth, *The Review of Economics and Statistics*, 91(1): 66-82

[9] Bils, Mark and Peter Klenow(2000), "Does Schooling Cause Growth?", *American Economic Review*, 90:1160-1183,

[10] De la Feunte, Angel and Rafael Domenech(2002), "Human Capital in Growth Regressions: How Much Difference does Data Quality Make? An Update and Further Results", *mimeo*, August 2002

[11] Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table", *American Economic Review*, 105(10): 3150-3182

[12] Ha, J., Kim, Y-J., and Lee, J-W., (2009) "Optimal Structure of Technology Adoption and Creation: Basic Versus Development Research in Relation to the Distance from the Technological Frontier", *Asian Economic Journal* 23(3):373-95

[13] Hanueshek, E. A. and Kimko, D. D. (2000), Schooling, Labor-force Quality, and the Growth of Nations, *American Economic Review*, 90(5): 1184-1208

[14] Islam, M. (2010), Human Capital Composition, Proximity to Technology Frontier and Productivity Growth, *Monach University Department of Economics, Discussion Paper*, 23(10): 1-41

[15] Kim, Y.J. and Lee, J-W. (2011), A Survey on the Relationship between Education and Growth with Implications for Developing Asia, *Journal of International Commerce, Economics and Policy*, 4(01):135005.

[6] Kreuger, A.B., and Lindahl, M. (2001), Education for Growth: Why and for Whom? *Journal of Economic Literature*, 39(4): 1101-1136

[17] Mankiw, N. G., Romer, D., and Weil, D.N. (1992), A Contribution to the Empirics of Economic Growth, *The Quarterly Journal of Economics*, 107(2): 407-437.

[18] Mincer, J. (1958), Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, 66(4): 281-302

[19] Nelson, R.R. and Phelps, E.S. (1966), Investment in Humans, Technological Diffusion and Economic Growth, *The American Economic Review*, 56(1/2):69-75

[20] OECD, "Education at a Glance", 2002

[21] Vandenbussche, J., Aghion, P., and Meghir, C. (2006), Growth, Distance to Frontier and Composition of Human Capital, *Journal of Economic Growth*, 11(2): 97-127

초 록

이 논문에서는 고등교육과 기술 성장 사이의 관계를 규명하는 실증 연구의 잠재적인 문제점을 지적한다. 구체적으로 논문에서는, 기존 연구의 결과에 따라, 고등교육이 기술적으로 진보한 나라에서 더 중요한 역할을 할 것이라는 가정을 하고, 이를 규명하기 위해 각 국가의 고등교육 이수자 비율, 그리고 해당 국가가 기술적으로 진보한 정도, 두 변수의 상호 작용을 (interaction term) 회귀분석에 포함시켜 실증분석을 시도하였다. 하지만, 이에 따른 분석결과가 패널에 포함된 국가와 분석기간(timeline)이 달라지면 기존 연구와 결과가 상이해진다는 점을 확인할 수 있었다. 특히, 기존의 실증 연구에서 확인되었던 고등교육과 기술진보 사이의 양(+)의 관계가, 해당 국가가 급속성장하는 기간에는 더 이상 성립하지 않는 다는 점도 확인되었다. 결론적으로, 교육과 기술진보 사이의 보편적인 관계가 규명되려면, 기존 모델에 대한 추가적인 연구가 선행되어야 한다.

주요어 : 대학교육, 기술진보, 인적 자본의 구성

학 번 : 2016-29711