Black Gold and Dull Minds: The Impact of Hydrocarbon Exploration Announcements on Education in Colombia *

Polanco-Jiménez, Jaime
Pontifical Xavierian University
jaime.polanco@javeriana.edu.co

Christian Posso
Banco de la República de Colombia
`cpossosu@banrep.gov.co`

May 9, 2023

Abstract

This study investigates the impact of oil exploration announcements on dropout rates of students in the exploration areas of Colombia, with a particular focus on the gender heterogeneity effect. To identify the causal effect, we employ a Geographic Regression Discontinuity Design. The findings reveal a significant reduction in the dropout rate of elementary school pupils, with a stronger effect for girls than for boys. The results suggest that the impact of natural resources on education outcomes is a crucial issue for policymakers, and measures need to be taken to ensure that these effects are considered in the development of educational policies in the exploration areas, particularly for girls.

JEL Codes: I21, O13.

Key Words: oil and gas exploration, education, human capital accumulation, labor force,

*I thank Oliver Pardo for his relentless supervision and support. Thanks to Brigitte Castaneda, Gloria Bernal and Luz Abadia for their suggestions, support and for providing valuable elements that allowed this project to be carried out.

1 Introduction

The relationship between oil industry development and education has been a subject of interest for policymakers, scholars, and researchers alike. While oil revenues have the potential to fund education and increase human capital, the focus on oil extraction may also lead to the neglect of other sectors, including education. Therefore, it is crucial to understand the impact of oil and gas exploration announcements under a concession agreement of exploration and production (E&P) in all their stages on educational outcomes, particularly in developing countries like Colombia.

In the context of investigating the impact of oil and gas exploration announcements on youth dropout rates and education performance in Colombia, it is important to note that such announcements typically generate an increase in expected income. The announcement of oil exploration occurs within 90 days after the signing of the first stage of exploration, known as the "Preliminary Phase" or "Phase 0". During this phase, geophysical and geological studies are conducted, as well as stages of verification, confirmation, and/or certification of possible presence of ethnic groups or communities in the area of influence. If such communities exist, prior consultation activities are undertaken. However, this phase does not involve actual exploration work or investments in infrastructure, and therefore, it only includes speculations on expected future income.

The National Hydrocarbons Agency (ANH) is the Colombian government agency responsible for identifying and regulating areas with hydrocarbon potential, where the location of hydrocarbon deposits is determined solely based on geological. By analyzing the effects of ANH concessions on educational outcomes, particularly schooling decisions and education performance, this study aims to contribute to the literature on the relationship between oil industry development and education.

The literature on oil and gas industry highlights a mixed impact on education. While the presence of oil fields may negatively affect education (Farzanegan and Thum (2018) , Marchand and Weber (2018) , (Farzanegan, 2017) , Genareo (2018) , Zuo, Schieffer, and Buck (2019)), oil revenues have the potential to positively impact education by providing funding for improving its quality and promoting human capital development (Fábio Bentz Maciel (2021) , Kumar (2017)). However, the effectiveness of this impact is contingent upon the allocation and management of funds.

To demonstrate a causal effect of exploration announcements on educational outcomes in this study we use a Geographic Regression Discontinuity Design (GRD) as a natural experiment design. On average, students in schools near the border of the exploration area under the exploration and production (E&P) contract have similar sociodemographic conditions, and there is no migration of students due to the exploration announcement. Furthermore, the geographical conditions on both sides of the border are similar.

The results of article suggest that income expectations of those who live within the exploration area increase with the announcement, which results in a modification of their present behavior, valuing the accumulation of human capital for future time. This is reflected in a decrease in student desertion, a decrease in the number of students in the labor force of students in the last grade of high school education, an increase in the results in standardized tests of students in mathematics and language. These findings provide valuable insights for policymakers and stakeholders working to mitigate any negative effects of the oil and gas industry on education in Colombia.

The article is organized as follows: Section 2 provides a comprehensive review of the literature on the impact of hydrocarbon fields on education. Section 4 presents the institutional context surrounding oil and gas exploration and production concessions in Colombia. Section 5 discusses the conceptual framework of the study, while Section 6 presents the mechanisms by which a hydrocarbon field can affect education. Section 3 describes the data used in the study, while Section 7 presents the results. Finally, Section 8 draws conclusions.

2 Literature Review

The oil industry has played a significant role in shaping the economies of many countries, particularly those that are heavily dependent on oil revenues. As a result, there have been numerous studies examining the impact of oil rents on various aspects of economic development, including education. In this study, we will summarize the findings of several studies that have investigated the schooling decision and academic performance and the relationship with oil and gas industry.

Literature Review: The motive to education are: individuals attend school for its expected future higher-wage reward (Juntip 2006) the joy of learning new things, meeting new people (Juntip 2006), the ability to consume a wider range of goods in life (Juntip 2006). Induced social status individuals can get through education (Juntip 2006) Milton Friedman argued that giving parents freedom to choose schools would improve education (MacLeod 2019) households often seem to choose schools based on their absolute achievement rather than their value added education is an investment into human capital. labor markets can feature wage premia: Individuals of a given skill level may receive higher wages if they match to more productive firms distance influences school choice and the placements that schools produce.

Literature Review: Do changes in income expectations lead to changes in behavior? Households update their income forecast and adjust consumption plans accordingly. (Das 1997, Roth 2017 and Roth 2018) found that households underestimate their future incomes, Jappelli 2009: Further research is needed to more fully understand the relationship between income expectations and behavior. Armantier 2011: Survey respondents act on inflation expectations, irrationals have lower education and financial literacy. Das 1999: Respondents form rational expectations, reported expectations are best future predictions.

Literature Review: What is the effect of income changes on academic performance? Duncan 2011: A \$1,000 increase in annual income raises young children's achievement by 5 %-6 % of a standard deviation. Dahl 2008: A \$1,000 increase in income raises math and reading test scores by 6 % of a standard deviation in the short-run. Chmielewski 2016: US income achievement gap larger than other countries. Carlisle 2015: Socioeconomic segregation, school funding, teacher expectations, and academic climate affect academic achievement

The study conducted by Farzanegan (2017) found that oil-dependent economies tend to show a lower economic growth rate compared to resource-poor countries. They also found that oil-rich countries still suffer from an insufficient quality of primary and secondary education. This is likely due to the negative impact that oil rents can have on education by diverting resources away from education and into other areas.

Another study by Marchand and Weber (2018) Weber found that resource development can slightly decrease student achievement, even though it provides schools with more money. This was demonstrated in their intervention examining the effect of the Texas shale boom on teacher quality and student achievement. The outcomes measured were student achievement, teacher quality, and student productivity. The researchers found that the growing gap in wages between the private and education sectors contributed to greater teacher turnover and more inexperienced teachers, which in turn negatively impacted student achievement.

In contrast, Farzanegan and Thum (2018) found that oil-induced employment opportunities for natives can increase secondary school attendance in the region. Their intervention studied the effect of oil induced employment opportunities for natives on secondary schooling decisions in the oil rich region of Chad. The outcomes measured were secondary school attendance and school dropouts. The study found that oil-induced employment opportunities increased secondary school attendance in the region, indicating that the presence of the oil industry can have a positive impact on education.

Genareo (2018) found that teachers faced three key challenges in their classrooms as a result of the population influx. These challenges included changing educational space, student academic proficiency, and a lack of cultural competence and pedagogical knowledge to effectively educate new, diverse students. This highlights the importance of considering the impact of oil rents on the quality of education and the need for adequate resources and support for teachers in oil-dependent economies.

Zuo et al. (2019) found that intensive drilling activities decreased grade 11 and 12 enrollment over the 14-year study period. Their intervention examined the effect of the oil and gas boom on schooling decisions in the U.S. The outcome measured was grade 11 and 12 enrollment. The study highlights the potential negative impact that the oil industry can have on education and the need to address these impacts in order to ensure that students have access to quality education.

Erdoğan, Yıldırım, and Gedikli (2020) found that there was no strong awareness of the importance of education in the countries included in the analysis. Their intervention studied the relationship between oil revenues and education. The outcomes measured were the total number of students enrolled in postsecondary education, general programs, and general and private high school education institutions. This highlights the need for increased awareness and investment in education in oil-dependent economies.

Chuan (2022) examines the effect of oil and gas job opportunities during youth on human capital and college enrollment. Using an instrumental variable approach, the author finds that exposure to oil and gas job opportunities during youth decreases college-going rates for men, but not women. This leads to permanent declines in college attainment, but gains in employment and earnings at ages 25-30. The author posits that these gains are driven by cohorts who reach college age during industry booms, suggesting that informal human capital can compensate for the loss of schooling for men who leave school for oil and gas work. The outcomes measured in the paper are educational attainment, employment, and earnings.

Birks and Rimmer (1984) found that the speed of change in the states of Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates makes analysis of educational development especially interesting in itself. This highlights the importance of ongoing research and analysis in order to understand the impact of oil rents on education and the need for appropriate interventions and

Kumar (2017) found that the oil boom of the 1970s had an impact on human capital investment through two channels, the growth in the relative demand for skills in the oil and gas sector and the real wage premium for a college degree. Charles and Ngozi-Ohehi (2018) found that multinational oil companies should employ teachers who will teach in public schools located in oil-rich areas. M. Moustapha (2021) found that oil-induced employment opportunities for natives increased secondary school attendance in Chad.

This study aims to contribute to the existing literature on the impact of natural resource abundance on human capital accumulation. Specifically, the study focuses on the effect of oil exploration announcements on educational outcomes, including schooling decisions, academic performance, labor force participation, and student mobility. Furthermore, the study examines the heterogeneous on gender.

To estimate the effect of oil exploration on educational outcomes, the study developed a software framework in Python that uses geographic regression discontinuity design to estimate euclidean distance, real distance by walking, and driving distance. The framework takes a polygon as input to define units within the area under treatment and control, and georeferenced units of observation to estimate the effect of being within the area of treatment A.1. The study's findings have significant implications for policymakers and can inform decisions related to natural resource management and education policy.

3 Data

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

3.1 Schools

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

3.2 Academic trajectory

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

3.3 Student Caracteristics

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

3.4 Areas under an E&P Contract

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

Graph/The oil field blessing.png

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

3.5 Academic performance

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

4 Oil and Gas exploration and production contracts in Colombia.

The exploration and production contract (E&P) in Colombia is a legal agreement between the National Hydrocarbons Agency (ANH) and a company for the exploration and exploitation of hydrocarbons in a given area.

The National Hydrocarbons Agency (ANH) is a Colombian government agency tasked with the management and regulation of the nation's hydrocarbon resources. As such, the ANH is responsible for identifying areas that have the potential for hydrocarbon deposits through a rigorous and methodical assessment process.

In order to identify Exploration and Production (E&P) areas that possess hydrocarbon potential, the ANH employs a variety of techniques and methodologies that take into account various factors such as the geological and mining characteristics of the rocks, as well as the results of previous exploration efforts in similar geologic strata. This process involves a systematic and comprehensive evaluation of the geologic conditions in order to identify areas that have a higher likelihood of containing hydrocarbons.

- The contract can be awarded through a public bidding process or direct negotiation.
- The contract includes the technical and economic requirements for the exploration and production activities, as well as the rights and obligations of the parties involved.
- The contract also establishes the royalties, taxes, and other payments that the company must make to the government, as well as the social and environmental obligations.

- The exploration phase can last up to 7 years and includes activities such as seismic surveys, drilling, and well testing.
- The production phase can last up to 30 years and includes the actual extraction of hydrocarbons, as well as the transportation and commercialization of the products.
- The contract can be extended for additional periods, subject to the fulfillment of certain conditions.

It is of paramount importance to note that the location of hydrocarbon deposits is determined solely on the basis of geological conditions, without any consideration given to their proximity to educational institutions or other non-geological factors. This ensures that the areas selected for exploration and production are those that possess the greatest hydrocarbon potential, and that the assessment process is fair and objective.

Furthermore, the process of Exploration and Production (E&P) under a concession contract, called the E&P contract, as this is a special state contract for the exploration and exploitation of non-renewable natural resources. The ANH grants the contractor the right to carry out, at his own risk and expense, the exploration and exploitation activities of the area granted by concession to the state.

Registros historicos del programa. Cuantas zonas than sido intervenidas. https://www.anh.gov.co/es/hidrocarburos/oportunidades-disponibles/

5 Conceptual framework

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

6 Mechanisms

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx.

6.1 Empirical Strategy

In order to identify the causal effect of hydrocarbon exploration announcements on dropout rates and academic performance of students in the exploration areas of Colombia, we employ a Geographic Regression Discontinuity Design. In this study, the educational institutions treated are those that were influenced by an Exploration and Production (E&P) concession that, during its exploration phase, resulted in hydrocarbon findings that led to the opening of a gas or oil field. The control educational institutions are those that were influenced by an E&P concession but in their exploration phase, did not find hydrocarbon discoveries or their discovery does not lead to economically viable exploitation. Our research will measure the educational outcome and compare the two groups and thus demonstrate the causality.

Initially, we will use ordinary least squares regression to estimate the impact of the oil field opening on the academic performance of students (i) who attended schools (s)

within a radius of one kilometer from the oil field. This will allow us to determine any correlations between the proximity of the oil field and the educational outcomes of the students.

$$Y_{s,i,T} = \alpha + \beta_1 \cdot oil_{field_T} + \varepsilon_{s,i} \tag{1}$$

Where $Y_{s,i,T}$ is the outcome on education of student i in the school s in the period relative to the aperture of oil field T.

7 Results

Results for Grade 1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2135	0.036	6.003	0.000	0.143	0.284
treated by the announcement of exploration	-0.1304	0.047	-2.751	0.008	-0.225	-0.036

Results for Grade 2

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2445	0.040	6.060	0.000	0.164	0.325
treated by the announcement of exploration	-0.1588	0.054	-2.965	0.004	-0.266	-0.052

Results for Grade 3

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.1944	0.042	4.627	0.000	0.111	0.278
treated by the announcement of exploration	-0.1017	0.056	-1.821	0.073	-0.213	0.010

Results for Grade 4

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2168	0.047	4.656	0.000	0.124	0.310
treated by the announcement of exploration	-0.0843	0.061	-1.374	0.174	-0.207	0.038

Results for Grade 5

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2978	0.047	6.347	0.000	0.204	0.391
treated by the announcement of exploration	-0.1376	0.064	-2.163	0.034	-0.265	-0.011

Results for Grade 6

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.1278	0.057	2.223	0.037	0.009	0.247
treated by the announcement of exploration	0.1112	0.070	1.581	0.128	-0.035	0.257

Results for Grade 7

	coef	std err	t	P> |t|	[0.025	0.975]
const	0.1508	0.076	1.995	0.059	-0.006	0.308
treated by the announcement of exploration	0.0659	0.093	0.712	0.484	-0.126	0.258

Results for Grade 8

	coef	std err	t	P> |t|	[0.025	0.975]
const	0.1582	0.077	2.041	0.053	-0.003	0.319
treated by the announcement of exploration	0.0834	0.095	0.879	0.389	-0.113	0.280

Results for Grade 9

	coef	std err	t	P> |t|	[0.025	0.975]
const	0.1488	0.092	1.621	0.120	-0.042	0.340
treated by the announcement of exploration	0.1725	0.114	1.518	0.144	-0.064	0.409

Results for Grade 10

	coef	std err	t	P> |t|	[0.025	0.975]
const	0.2232	0.111	2.004	0.060	-0.011	0.457
treated by the announcement of exploration	0.0846	0.133	0.635	0.533	-0.195	0.364

Results for Grade 11

	coef	std err	t	P> |t|	[0.025	0.975]
const	0.9913	0.021	48.169	0.000	0.948	1.035
treated by the announcement of exploration	-0.0508	0.025	-2.016	0.061	-0.104	0.003

Table 1: General result at the time the oil field has opened.

	Dependent variable:			
	Number of Students	frac. Students in labor force	Math Score	Reading Score
	(1)	(2)	(3)	(4)
Treated	8.523	-0.063^{***}	-1.524^{**}	-1.326^{***}
	(5.233)	(0.018)	(0.607)	(0.503)
Constant	66.744^{***}	0.239^{***}	46.735^{***}	48.019^{***}
	(2.083)	(0.007)	(0.242)	(0.200)
Observations	1,016	1,016	1,016	1,016
R^2	0.003	0.012	0.006	0.007
Adjusted R^2	0.002	0.011	0.005	0.006
Residual Std. Error (df = 1014)	60.915	0.210	7.066	5.852
F Statistic (df = 1; 1014)	2.653	12.072^{***}	6.302^{**}	6.957^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2: Math score in the stage of exploration l

	Dependent variable:			
	Math Score			
	2y after signed	4y after signed	6y after signed	8y after signed
TREAT_1	-1.551^{***}	-1.251^{**}	-1.235^{**}	-0.299
	(0.598)	(0.531)	(0.549)	(0.651)
Constant	45.856^{***}	45.719^{***}	46.549^{***}	47.001^{***}
	(0.196)	(0.175)	(0.210)	(0.246)
Observations	1,153	1,413	1,092	1,040
R^2	0.006	0.004	0.005	0.0002
Adjusted R^2	0.005	0.003	0.004	-0.001
Residual Std. Error	6.290 (df = 1151)	6.215 (df = 1411)	6.410 (df = 1090)	7.350 (df = 1038)
F Statistic	6.729^{***} (df = 1; 1151)	5.558^{**} (df = 1; 1411)	5.071^{**} (df = 1; 1090)	0.212 (df = 1; 1038)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3: General results during first stage of production

	Dependent variable:				
	Math Score				
	1y in production	3y in production	5y in production	7y in production	9y in production
TREAT_1	-0.299	-0.377	-0.526	2.310^{**}	-0.996
	(0.651)	(0.608)	(0.510)	(1.019)	(0.856)
Constant	47.001^{***}	47.707^{***}	46.345^{***}	44.736^{***}	44.859^{***}
	(0.246)	(0.234)	(0.207)	(0.684)	
Observations	1,040	1,249	1,457	220	239
R^2	0.0002	0.0003	0.001	0.023	0.006
Adjusted R^2	-0.001	-0.0005	0.00004	0.019	0.001
Residual Std. Error	7.350 (df = 1038)	7.638 (df = 1247)	7.229 (df = 1455)	7.520 (df = 218)	6.367 (df = 237)
F Statistic	0.212 (df = 1; 1038)	0.384 (df = 1; 1247)	1.064 (df = 1; 1455)	5.137^{**} (df = 1; 218)	1.355 (df = 1; 237)
Note:				*p<0.1; **p<0.05; ***p<0.01	

https://www.sciencedirect.com/science/article/pii/S0304387820301565 Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.??

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx. (?)

Table 4: General results by period after exploration stage

	Dependent variable:			
	Number of Students	frac. Students in labor force	Math Score	Reading Score
	(1)	(2)	(3)	(4)
treat	8.523	-0.063^{***}	-1.524^{**}	-1.326^{**}
	(6.143)	(0.016)	(0.632)	(0.532)
Period :8	2.633	0.055^{***}	0.265	0.628^{**}
	(3.423)	(0.009)	(0.352)	(0.297)
Period :9	12.617^{***}	0.096^{***}	-0.336	0.305
	(3.053)	(0.008)	(0.314)	(0.265)
Period :10	2.670	0.142^{***}	0.972^{***}	1.521^{***}
	(3.284)	(0.008)	(0.338)	(0.285)
Period :11	26.478^{***}	0.108^{***}	0.198	0.462
	(3.503)	(0.009)	(0.360)	(0.303)
Period :12	19.247^{***}	0.087^{***}	-0.390	0.073
	(3.191)	(0.008)	(0.328)	(0.277)
Period :13	-5.845	0.164^{***}	0.324	0.972^{***}
	(3.922)	(0.010)	(0.403)	(0.340)
Period :14	4.719	0.091^{***}	-2.000^{***}	-1.159^{*}
	(6.945)	(0.018)	(0.714)	(0.602)
Period :15	0.256	0.152^{***}	-0.782	-0.250
	(7.733)	(0.020)	(0.795)	(0.670)
Period :16	55.986^{***}	0.017	-1.877^{***}	-0.742
	(6.294)	(0.016)	(0.647)	(0.545)
Period :17	21.743^{**}	0.080^{***}	-0.005	1.353^{*}
	(8.665)	(0.022)	(0.891)	(0.751)
$trear \cdot Period$:8	-1.988	-0.082^{***}	1.225	0.660
	(8.820)	(0.023)	(0.907)	(0.764)
$trear \cdot Period$:9	15.078^{*}	-0.030	0.642	0.617
	(7.789)	(0.020)	(0.801)	(0.675)
$trear \cdot Period$:10	-0.618	-0.017	1.147	0.498
	(8.378)	(0.022)	(0.861)	(0.726)
$trear \cdot Period$:11	-26.126^{***}	-0.017	1.686^{*}	1.706^{**}
	(9.198)	(0.024)	(0.946)	(0.797)
$trear \cdot Period$:12	-2.303	0.005	0.998	0.892
	(7.947)	(0.020)	(0.817)	(0.689)
$trear \cdot Period$:13	12.510	-0.011	0.718	1.033
	(8.300)	(0.021)	(0.853)	(0.719)
$trear \cdot Period$:14	-18.622	0.135^{***}	3.834^{***}	3.330^{***}
	(11.474)	(0.029)	(1.179)	(0.994)
$trear \cdot Period$:15	5.657	0.063^{*}	1.831	1.228
	(13.922)	(0.036)	(1.431)	(1.206)
$trear \cdot Period$:16	-15.242	0.117^{***}	0.527	0.079
	(11.408)	(0.029)	(1.173)	(0.988)
$trear \cdot Period$:17	-34.875^{**}	0.115^{***}	0.389	-0.032
	(15.654)	(0.040)	(1.609)	(1.356)
Constant	66.744^{***}	0.239^{***}	46.735^{***}	48.019^{***}
	(2.445)	(0.006)	(0.251)	(0.212)
Observations	8,986	8,986	8,986	8,986
R^2	0.029	0.083	0.009	0.011
Adjusted R^2	0.027	0.080	0.007	0.009
Residual Std. Error (df = 8964)	71.507	0.184	7.351	6.195
F Statistic (df = 21; 8964)	12.922^{***}	38.435^{***}	3.845^{***}	4.793^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Average sample description (2010-2019)

Insurers	Providers	Relationships	States	Cities
37	1579	6737	27	632

Note: The table reports the average number of providers, insurers, relationships between them, states and cities included in the main dataset between 2010 and 2019.

7.1 The effect xxx

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

7.1.1 Main results

Graph/SA_Math_long.png

Graph/SA_Math_long.png

Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxx xxxxxxxxx.

Figure 1: First Stage using National aggregated data at insurance company Level

Figures/comparison Minsalud.pdf

Notes: Each dot represents an insurer that is receiving new affiliates in a given bankruptcy event, and the total number of affiliates is taken at the national level. The red line fits these dots using HC3 robust standard errors. The coefficient and standard error of this regression are shown in the figure.

8 Conclusion

This paper evaluates the effect Xxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxxxxxxxxxxx xxxxxxxxx

In conclusion, the research shows that while the oil industry can have both positive and negative impacts on education, it is important to consider the broader societal context and to implement policies that maximize the positive effects and minimize the negative ones. For example, policies that increase the quality of education, such as teacher training programs, can mitigate the negative effects of oil-based economies. Furthermore, it is crucial for multinational oil companies to invest in education and human capital development in oil-rich areas, as it will benefit not only the individuals but also the broader society. Ultimately, the research highlights the need for a nuanced and comprehensive approach to understanding the relationship between oil and education.

References

Birks, J., & Rimmer, J. (1984, jan 1). Developing Education Systems in the Oil States of Arabia: Conflicts of Purpose and Focus. *International Journal of Manpower*, *5*(1), 13–23. Retrieved from http://dx.doi.org/10.1108/EB044952 doi: 10.1108/eb044952

Charles, E. C., & Ngozi-Ohehi, L. C. (2018, jun 30). Teacher's Perception Of The Contributions Of Oil Exploration To Educational Development In Oguta L.G.A. *Advances in Social Sciences Research Journal*. Retrieved from http://dx.doi.org/10.14738/ASSRJ.56.4795 doi: 10.14738/assrj.56.4795

Chuan, A. (2022). The impact of oil and gas job opportunities during youth on human capital.
doi: http://dx.doi.org/10.2139/ssrn.3597176

Erdoğan, S., Yıldırım, D. , & Gedikli, A. (2020, jan 1). Relationship BETWEEN OIL REVENUES AND EDUCATION IN GULF COOPERATION COUNCIL COUNTRIES. *International Journal of Energy Economics and Policy*, *10*(1), 193–201. Retrieved from http://dx.doi.org/10.32479/ijeep.8653 doi: 10.32479/ijeep.8653

Farzanegan, M. R. (2017). More Oil, Less Quality of Education? New Empirical Evidence. *Center of Public and International Economics MORE*(July). doi: 10.13140/RG.2.2.36794.49608

Farzanegan, M. R., & Thum, M. (2018, sep 3). Does oil rents dependency reduce the quality of education? *Empirical Economics*, *58*(4), 1863–1911. Retrieved from http://dx.doi.org/10.1007/S00181-018-1548-Y doi: 10.1007/s00181-018-1548-y

Fábio Bentz Maciel. (2021). *Oil Windfalls and Educational Development.*

Genareo, V. (2018, nov 12). Policies and Professional Development: An Oil Boom's Effect on Rural Schools and Teachers. *The Rural Educator*, *37*(2). Retrieved from http://dx.doi.org/10.35608/RURALED.V37I2.268 doi: 10.35608/ruraled.v37i2.268

Kumar, A. (2017, jan 12). Impact of oil booms and busts on human capital investment in the USA. *Empirical Economics*, *52*(3), 1089–1114. Retrieved from http://dx.doi.org/10.1007/S00181-016-1192-3 doi: 10.1007/s00181-016-1192-3

M. Moustapha. (2021). *Oil boom, job prospects and schooling decisions : Evidence from Chad.*

Marchand, J., & Weber, J. (2018). The Local Effects of the Texas Shale Boom on Schools, Students, and Teachers. *SSRN Electronic Journal*. Retrieved from http://dx.doi.org/10.2139/ssrn.3096293 doi: 10.2139/ssrn.3096293

Zuo, N., Schieffer, J., & Buck, S. (2019, 2). The effect of the oil and gas boom on schooling decisions in the U.S. *Resource and Energy Economics*, *55*, 1–23. Retrieved from http://dx.doi.org/10.1016/J.RESENEECO.2018.10.002 doi: 10.1016/j.reseneeco.2018.10.002

A Appendix

A.1 Software: distancegeord in python language

```
!pip install distancegeord
from distancegeord import *
import pandas as pd
shape_polygon_path  ="./area_treatment.zip"
shape_units_path  ="./observation.zip"
data = distancegeord(shape_polygon_path, shape_units_path)
```

Figure A.1: Timeline of insurer liquidations

Figures/time_line.PNG

Source: own elaboration with information from *Supersalud*.

B Theorical Model

B.1 Model base: Influence of Family Background on Student Outcomes Varies by Grade Level

We will compare in base a theoretical microeconomic model of an education production function under the next assumption:

- The model takes into account the student's family attributes F, the student's attributes S, and the educational institution's attributes I.
- The outcome of the production function is the dropout rate.
- The first case is when there is no a Stronger Relationship Between Family and Student Attributes in Elementary School $\text{cov}(F_{sec}, S_{sec}) \simeq \text{cov}(F_{elem}, S_{elem})$.

Based on our model, we found that an increase in family attributes has a larger effect on the dropout rate in elementary school compared to secondary school. This is due to the fact that the covariance between family attributes and student attributes is larger in elementary school than in secondary school.

One possible explanation for this is that students in elementary school are more dependent on their families for support and guidance compared to students in secondary school. This could manifest in several ways - for example, younger children may require more assistance with homework, and may rely more heavily on their parents to help them stay motivated and engaged with their studies. As a result, changes in family attributes (such as an increase in income or parental education) could have a larger impact on the academic performance and dropout rates of elementary school students.

Another possible explanation is that families may play a different role in shaping educational outcomes at different stages of a student's academic career. For example, in secondary school, students may be more influenced by peer groups, teachers, and other external factors. In contrast, in elementary school, families may play a larger role in shaping a child's attitudes and behaviors towards education. As a result, changes in family attributes may have a larger impact on the academic outcomes of elementary school students.

Overall, the larger effect of family attributes on elementary school dropout rates highlights the importance of considering the role of families in shaping educational outcomes. By understanding the complex interplay between family and student attributes, educational policymakers can develop more effective strategies for promoting academic success and reducing dropout rates.

B.1.1 No Stronger Relationship Between Family and Student Attributes in Elementary School $\mathbf{cov}(F_{sec}, S_{sec}) \simeq \mathbf{cov}(F_{elem}, S_{elem})$.

In this case, students are equally dependent on their families in both elementary and secondary schools.

Assuming a Cobb-Douglas production function, the theoretical microeconomic model of an education production function can be represented as follows:

The production function is given by:

$$D = AF^{\alpha}S^{\beta}I^{\gamma}\epsilon, \tag{2}$$

where D is the dropout rate, F represents the student's family attributes, S represents the student's attributes, I represents the institutional attributes, A is a positive constant, α, β, and γ are positive constants that represent the elasticity of dropout rate with respect to each input, and ϵ is a random variable that captures all other factors affecting dropout rate that are not captured by the inputs.

The producer's problem is to maximize the production function, subject to the budget constraint:

$$C = w_F F + w_S S + w_I I, \tag{3}$$

where C is the total cost of inputs, w_F, w_S, and w_I are the prices of each input.

The Lagrangian is:

$$\mathcal{L} = AF^{\alpha}S^{\beta}I^{\gamma}\epsilon - \lambda(w_F F + w_S S + w_I I - C), \tag{4}$$

where λ is the Lagrange multiplier.
Taking partial derivatives and equating them to zero, we get:

$$\frac{\partial \mathcal{L}}{\partial F} = \alpha A F^{\alpha-1} S^{\beta} I^{\gamma} \epsilon - \lambda w_F = 0, \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial S} = \beta A F^{\alpha} S^{\beta-1} I^{\gamma} \epsilon - \lambda w_S = 0, \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial I} = \gamma A F^{\alpha} S^{\beta} I^{\gamma-1} \epsilon - \lambda w_I = 0. \tag{7}$$

Solving for F, S, and I, we get:

$$F = \left(\frac{\alpha w_F}{A\epsilon}\right)^{\frac{1}{\alpha-1}} \cdot \left(\frac{S}{I}\right)^{\frac{\beta}{\alpha-1}} \tag{8}$$

$$S = \left(\frac{\beta w_S}{A\epsilon}\right)^{\frac{1}{\beta-1}} \cdot \left(\frac{F}{I}\right)^{\frac{\alpha}{\beta-1}} \tag{9}$$

$$I = \left(\frac{\gamma w_I}{A\epsilon}\right)^{\frac{1}{\gamma-1}} \cdot \left(\frac{F}{S}\right)^{\frac{\alpha+\beta}{\gamma-1}}. \tag{10}$$

The optimal level of production is given by:

$$D^* = A \left(\frac{\alpha w_F}{A\epsilon}\right)^{\frac{\alpha}{\alpha-1}} \left(\frac{\beta w_S}{A\epsilon}\right)^{\frac{\beta}{\beta-1}} \left(\frac{\gamma w_I}{A\epsilon}\right)^{\frac{\gamma}{\gamma-1}}. \tag{11}$$

The marginal cost of education is:

$$MC = \frac{\partial C}{\partial E} = w_S + w_F \frac{\partial S}{\partial E} + w_I \frac{\partial I}{\partial E}$$

where w_S, w_F, and w_I are the unit prices of the inputs S, F, and I, respectively.
The optimal level of education output for the elementary school is:

$$E_{elem} = \left(\frac{\alpha_S w_S}{w_F \text{cov}(F_{elem}, S_{elem})}\right)^{\frac{1}{1-\alpha}} I_{elem}^{\frac{\alpha_I}{1-\alpha}}$$

And the optimal level of education output for the secondary school is:

$$E_{sec} = \left(\frac{\alpha_S w_S}{w_F \text{cov}(F_{sec}, S_{sec})}\right)^{\frac{1}{1-\alpha}} I_{sec}^{\frac{\alpha_I}{1-\alpha}}$$

Now, let's assume that we have a program that increases the expected value of family attributes for both elementary and secondary schools, but does not affect the other inputs. Specifically, the program increases the expected value of family attributes from F_{elem} to $F_{elem} + \Delta F$ and from F_{sec} to $F_{sec} + \Delta F$. We want to evaluate the effect of this program on the dropout rate for both schools.

To do this, we first need to calculate the new optimal level of education output for each school. From our previous equations, we know that the optimal level of education output is a function of the inputs and the unit prices of those inputs. Since the program only affects F, the unit prices of S and I do not change. Therefore, the only thing that changes is the optimal level of education output for each school. The new optimal level of education output for the elementary school is:

$$E^*_{elem} = \left(\frac{\alpha_S w_S}{w_F \text{cov}(F_{elem} + \Delta F, S_{elem})}\right)^{\frac{1}{1-\alpha}} I_{elem}^{\frac{\alpha_I}{1-\alpha}}$$

And the new optimal level of education output for the secondary school is:

$$E^*_{sec} = \left(\frac{\alpha_S w_S}{w_F \text{cov}(F_{sec} + \Delta F, S_{sec})}\right)^{\frac{1}{1-\alpha}} I_{sec}^{\frac{\alpha_I}{1-\alpha}}$$

Now, we can calculate the new dropout rates for each school. Let D_{elem} and D_{sec} denote the original dropout rates for the elementary and secondary schools, respectively. The new dropout rate for the elementary school is:

$$D^*_{elem} = D_{elem} + \frac{\partial D_{elem}}{\partial E_{elem}}(E^*_{elem} - E_{elem})$$

where $\frac{\partial D_{elem}}{\partial E_{elem}}$ is the marginal effect of education on dropout rate for the elementary school. Similarly, the new dropout rate for the secondary school is:

$$D^*_{sec} = D_{sec} + \frac{\partial D_{sec}}{\partial E_{sec}}(E^*_{sec} - E_{sec})$$

The new dropout rate for the secondary school is:

$$D'_{sec} = D_{sec} - \frac{\partial D_{sec}}{\partial E_{sec}} \Delta E_{sec}$$

where $\frac{\partial D_{sec}}{\partial E_{sec}}$ is the marginal effect of education on dropout rate for the secondary school. Substituting in the values for D_{sec} and $\frac{\partial D_{sec}}{\partial E_{sec}}$ derived earlier, we have:

In it, we derived the expression for the new dropout rate for the secondary school after a small increase in the expected family attributes, given by:

$$D'_{sec} = 0.5 - 0.1F_{sec} - 0.1S_{sec}$$

Now, we want to evaluate the effect of an increase in the expected family attributes on the secondary school dropout rate. We know that $\frac{\partial D_{sec}}{\partial F_{sec}} = -0.1$, which means that a 1% increase in the expected family attributes leads to a 0.1% decrease in the secondary school dropout rate.

Assuming a small increase in expected family attributes of $\Delta F_{sec} = 0.1$, we can substitute this value into the expression above:

$$\begin{aligned} D'_{sec} &= 0.5 - 0.1(F_{sec} + \Delta F_{sec}) - 0.1S_{sec} \\ &= 0.5 - 0.1F_{sec} - 0.1\Delta F_{sec} - 0.1S_{sec} \\ &= 0.5 - 0.1F_{sec} - 0.01 - 0.1S_{sec} \\ &= 0.49 - 0.1F_{sec} - 0.1S_{sec} \end{aligned}$$

Thus, substituting in the values for D_{sec} and $\frac{\partial D_{sec}}{\partial E_{sec}}$ derived earlier, we get:

$$D'_{sec} = 0.5 - 0.1 \cdot 0.1 = 0.49$$

This means that a small increase in the expected family attributes of 0.1 leads to a 0.01 decrease in the secondary school dropout rate, assuming all other variables remain constant.

Thus, increasing the expected value of family attributes by one standard deviation reduces the secondary school dropout rate by approximately 0.01, or 1 percentage point.

Comparing the effect of the increase in family attributes on the dropout rates for the elementary and secondary schools, we see that the effect is larger for the elementary school. This is because the marginal effect of education on the elementary school dropout rate is larger than the marginal effect of education on the secondary school dropout rate, and the elementary school has a higher baseline dropout rate to begin with.

In summary, increasing the expected value of family attributes reduces the dropout rate for both elementary and secondary schools, but the effect is larger for the elementary school. This result is consistent with the assumption that the covariance between family attributes and student attributes is larger for the elementary school than for the secondary school.

B.1.2 Stronger Relationship Between Family and Student Attributes in Elementary School $\mathbf{cov}(F_{elem}, S_{elem}) \gg \mathbf{cov}(F_{sec}, S_{sec})$.

Let's assume that $\text{cov}(F_{elem}, S_{elem}) \gg \text{cov}(F_{sec}, S_{sec})$. We can then modify our production function as follows:

$$D_{elem} = f(S_{elem}, F_{elem}, I) + \varepsilon_{elem}$$

$$D_{sec} = f(S_{sec}, F_{sec}, I) + \varepsilon_{sec}$$

where ε_{elem} and ε_{sec} are error terms that capture the random variability in the dropout rate.

We can again assume that the production function takes the form of a Cobb-Douglas function as follows:

$$D_{elem} = A_{elem} S_{elem}^{\alpha_1} F_{elem}^{\alpha_2} I^{\alpha_3} e^{\varepsilon_{elem}}$$

$$D_{sec} = A_{sec} S_{sec}^{\beta_1} F_{sec}^{\beta_2} I^{\beta_3} e^{\varepsilon_{sec}}$$

where $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$, and β_3 are the elasticities of the dropout rate with respect to the inputs.

We want to model the educational institution's profit-maximizing behavior. The institution's revenue comes from the academic prestige and academic performance of students, and we can assume that the revenue is proportional to the dropout rate. So, we can write the institution's revenue as:

$$\text{Revenue} = P(D_{elem})D_{elem} + P(D_{sec})D_{sec}$$

where $P(D_{elem})$ and $P(D_{sec})$ are the prices (or tuition rates) per student for the elementary and secondary schools, respectively, and D_{elem} and D_{sec} are the corresponding dropout rates.

The institution's costs come from hiring teachers and staff, as well as investing in school and family attributes. We can assume that the prices of the inputs (i.e., wages and investments) are known and fixed, denoted by $w_{S_{elem}}, w_{S_{sec}}, w_{F_{elem}}$, and $w_{F_{sec}}$. Then, the institution's cost can be expressed as:

$$\text{Cost} = w_{S_{elem}} S_{elem} + w_{S_{sec}} S_{sec} + w_{F_{elem}} F_{elem} + w_{F_{sec}} F_{sec}$$

where S_{elem} and S_{sec} are the student attributes in elementary and secondary schools, respectively, and F_{elem} and F_{sec} are the corresponding family attributes.

where the maximization is subject to the production function and the prices of the inputs.

$$\max_{S_{elem}, F_{elem}, S_{sec}, F_{sec}} P(D_{elem}) D_{elem} + P(D_{sec}) D_{sec}$$
$$-\text{Cost}$$

where $P(D_{elem})$ and $P(D_{sec})$ are the prices of the dropout rate in the elementary and secondary schools, respectively, and $w_{S_{elem}}, w_{S_{sec}}, w_{F_{elem}}$, and $w_{F_{sec}}$ are the prices of student attributes, school attributes, family attributes in elementary school, and family attributes in secondary school, respectively.

Following the same steps as before, we can find the optimal levels of inputs for each school level. The optimal level of family attributes for each school level is given by:

$$F^*_{elem} = \frac{\alpha_2 P(D_{elem}) A_{elem}^{1/\alpha_1} I^{\alpha_3} w_{F_{elem}}}{w_{S_{elem}} \alpha_1 S_{elem}^{*(\alpha_1 - 1)}}$$

$$F^*_{sec} = \frac{\beta_2 P(D_{sec}) A_{sec}^{1/\beta_1} I^{\beta_3} w_{F_{sec}}}{w_{S_{sec}} \beta_1 S_{sec}^{*(\beta_1 - 1)}}$$

We can then substitute these expressions for F^*_{elem} and F^*_{sec} into the expressions for D_{elem} and D_{sec} and differentiate with respect to F_{elem} and F_{sec}, respectively, to obtain:

$$\frac{\partial D_{elem}}{\partial F_{elem}} = -\frac{\beta_{1,elem} \sigma_{F,S,elem}}{(\beta_{1,elem} \sigma_{F,S,elem} + \sigma_{S,elem})^2}$$

$$\frac{\partial D_{sec}}{\partial F_{sec}} = -\frac{\beta_{1,sec} \sigma_{F,S,sec}}{(\beta_{1,sec} \sigma_{F,S,sec} + \sigma_{S,sec})^2}$$

where $\sigma_{F,S,elem}$ and $\sigma_{F,S,sec}$ are the covariances between F and S for the elementary and secondary school, respectively.

If we assume that $\sigma_{F,S,elem} \ggg \sigma_{F,S,sec}$, then we can conclude that the effect of an increase in family attributes on dropout rates will be greater for the elementary school than for the secondary school. This is because a larger covariance between family and student attributes implies that family attributes are more important in determining student outcomes, and thus, an intervention that affects family attributes will have a larger impact on student outcomes.

To illustrate this point, suppose we assume that $\sigma_{F,S,elem} = 0.5$ and $\sigma_{F,S,sec} = 0.05$. Using the expressions for $\frac{\partial D_{elem}}{\partial F_{elem}}$ and $\frac{\partial D_{sec}}{\partial F_{sec}}$ derived earlier, we can compute the effect of a small increase in expected family attributes of $\Delta F_{elem} = \Delta F_{sec} = 0.1$ on the dropout rate for each school level:

$$\Delta D_{elem} = \frac{\partial D_{elem}}{\partial F_{elem}} \cdot \Delta F_{elem} = -0.091$$

$$\Delta D_{sec} = \frac{\partial D_{sec}}{\partial F_{sec}} \cdot \Delta F_{sec} = -0.009$$

We can see that the effect of the intervention on the elementary school is nearly ten times as large as the effect on the secondary school, as expected based on our assumption about the covariance between family and student attributes. This underscores the importance of considering the specific context in which an intervention is being implemented and how the relationships between different inputs may vary across different educational settings.

In summary, we have developed a theoretical microeconomic model of an education production function that takes into account the inputs of student and family attributes as well as institutional attributes. We have demonstrated how to calculate the optimal levels of inputs and the associated costs and benefits, as well as how to evaluate the effect of an intervention that increases the expected value of family attributes on dropout rates. We have also explored how the effect of such an intervention may differ between elementary and secondary schools based on differences in the covariance between family and student attributes.

B.2 General model

B.2.1 Heterogeneity in dropout rates between elementary and secondary school students.

Let us assume that the production of schools is now the dropout rate (D) instead of academic performance. We can still use the same functional form for the production function as before, but with dropout rate instead of academic performance:

$$D = f(E(F), S, I)$$

Assumption 1: Maximizes the production function when family attributes improve.

Assuming that an improvement in expected family attributes ($E(F)$) leads to a decrease in the dropout rate (D), we can state this assumption as:

$$\frac{\partial D}{\partial E(F)} < 0$$

This means that as expected family attributes improve, the dropout rate should decrease. We can demonstrate this assumption using a similar approach as before.

Let us assume that the production function for dropout rate is:

$$D = \alpha_1 E(F)^{\beta_1} S^{\beta_2} I^{\beta_3}$$

where $\alpha_1 > 0$, $\beta_1 < 0$, $\beta_2 < 0$, and $\beta_3 < 0$.

Taking the partial derivative of D with respect to $E(F)$, we get:

$$\frac{\partial D}{\partial E(F)} = \beta_1 \alpha_1 E(F)^{\beta_1 - 1} S^{\beta_2} I^{\beta_3}$$

Since $\beta_1 < 0$ and $\alpha_1 > 0$, we have $\frac{\partial D}{\partial E(F)} < 0$.

Assumption 2: Maximizes the production function when the expected value of family attributes improves.

Assuming that an improvement in the expected value of family attributes ($E(F)$) leads to a decrease in the dropout rate (D), we can state this assumption as:

$$\frac{\partial D}{\partial \mathbb{E}[F]} < 0$$

This means that as the expected value of family attributes improves, the dropout rate should decrease. To demonstrate this assumption, we can use the same approach as before.

Let us assume that the expected value of family attributes is given by:

$$\mathbb{E}[F] = \bar{F} + \epsilon_F$$

where $\bar{F}$ is the mean family attributes and ϵ_F is a random error term with mean zero.

We can then write the dropout rate as a function of the expected value of family attributes:

$$D = \alpha_1 (\bar{F} + \epsilon_F)^{\beta_1} S^{\beta_2} I^{\beta_3}$$

Taking the partial derivative of D with respect to the expected value of family attributes, we get:

$$\frac{\partial D}{\partial \mathbb{E}[F]} = \frac{\partial D}{\partial E(F)} \frac{\partial E(F)}{\partial \mathbb{E}[F]} = \beta_1 \alpha_1 (\bar{F} + \epsilon_F)^{\beta_1 - 1} S^{\beta_2} I^{\beta_3}$$

Since $\beta_1 < 0$, we have $\frac{\partial D}{\partial \mathbb{E}[F]} < 0$.

Now, let us explain why dropout rates for elementary and secondary school students may have different outputs. Dropout rates in elementary schools may depend more on family attributes, while dropout rates in secondary schools may depend more on individual attributes and school

Since $\beta_1 < 0$, we have $\frac{\partial D}{\partial \mathbb{E}[F]} < 0$. This means that an increase in expected family attributes leads to a decrease in the dropout rate, holding other factors constant. This is consistent with Assumption 2, which states that the production function is maximized when the expected value of family attributes improves.

Next, let's consider the dropout rate of students in secondary school. Let D_2 denote the dropout rate in secondary school, and let $E(F_2)$ denote the expected value of family attributes for students in secondary school. We can similarly specify the dropout rate production function as:

$$D_2 = f(E(F_2), S_2, I_2)$$

where S_2 represents the student's attributes in secondary school and I_2 represents the institutional attributes of the secondary school.

Again, we can use the total differential to express the change in the dropout rate as a result of a change in the expected value of family attributes:

$$dD_2 = \frac{\partial D_2}{\partial E(F_2)} dE(F_2) + \frac{\partial D_2}{\partial S_2} dS_2 + \frac{\partial D_2}{\partial I_2} dI_2$$

Using the same logic as before, we can see that an increase in expected family attributes will decrease the dropout rate in secondary school if:

$$\frac{\partial D_2}{\partial E(F_2)} < 0$$

This is consistent with Assumption 2, which states that the production function is maximized when the expected value of family attributes improves.

However, the effect of family attributes on the dropout rate in secondary school may be weaker than the effect on the dropout rate in elementary school. This is because students in secondary school may have more agency and make decisions about whether to continue their education based on factors beyond their family background. For example, they may drop out of school to work or for social reasons.

Therefore, we may have:

$$|\frac{\partial D_2}{\partial E(F_2)}| < |\frac{\partial D_1}{\partial E(F_1)}|$$

which means that a unit increase in expected family attributes may have a greater impact on the dropout rate in elementary school than on the dropout rate in secondary school.

In summary, the heterogeneity in the production function of dropout rates between elementary and secondary school arises because students in secondary school may have more agency and make decisions based on factors beyond their family background. However, both production functions are assumed to be decreasing in expected family attributes, consistent with Assumption 2.

B.3 General model with collorary

A microeconomic model of an education production function that take into account the student's family attributes F, the student's attributes St, and the educational institution's attributes I, where the educational outcome is the dropout rate.

Let D_{ij} denote the dropout rate for student i at school j. We can model the education production function as follows:

$$D_{ij} = f(F_i, St_i, I_j) + \epsilon_{ij}, \tag{12}$$

where $f(\cdot)$ is the production function, ϵ_{ij} is a random error term, F_i are the family attributes of student i, St_i are the student attributes, and I_j are the educational institution attributes.

B.3.1 Corollary 1: The effect of family attributes of students in elementary school on the dropout rate

The effect of family attributes of students in elementary school on the dropout rate, when student attributes don't change, can be represented by the partial derivative of the production function with respect to family attributes:

$$\frac{\partial D_{ij}}{\partial F_i} = \frac{\partial f(F_i, St_i, I_j)}{\partial F_i}. \tag{13}$$

2. The effect of expected family attributes of students in elementary school on the dropout rate, when student attributes don't change, can be represented by the expected partial derivative of the production function with respect to family attributes:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{14}$$

3. The difference between 1 and 2 can be represented by the expected partial derivative of the production function with respect to family attributes conditional on the expected student attributes:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{15}$$

These equations capture the effect of family attributes on the dropout rate in elementary school, both when considering individual differences in student attributes and when holding student attributes constant at their expected values. The difference between the two effects captures the interaction between family attributes and student attributes.

Demostration

Assuming that an increase in family attributes and expected family attributes have a negative effect on the dropout rate, we would expect that an increase in these attributes would lead to a decrease in the expected dropout rate.

To demonstrate this, let's first assume that the education production function is maximized with respect to all inputs, including family attributes. That is:

$$\max_{F_i, St_i, I_j} f(F_i, St_i, I_j). \tag{16}$$

Under this assumption, we can say that the expected dropout rate is at its lowest point, given the characteristics of the students and educational institutions.

Now, let's consider the effect of family attributes on the dropout rate. We can use Corollary 2 from the previous answer to demonstrate this effect:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{17}$$

Since we assume that an increase in family attributes has a negative effect on the dropout rate, we can conclude that the partial derivative of the production function with respect to family attributes is negative:

$$\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} < 0. \tag{18}$$

Therefore, we have:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] < 0. \tag{19}$$

This means that an increase in expected family attributes would lead to a decrease in the expected dropout rate.

In summary, if we assume that the education production function is maximized with respect to all inputs and that an increase in family attributes has a negative effect on the dropout rate, then an increase in expected family attributes would lead to a decrease in the expected dropout rate. This effect is captured by the expected partial derivative of the production function with respect to family attributes, which is negative.

B.3.2 Corollary 2: The effect of family attributes of students in secondary school on the dropout rate

The effect of family attributes of students in secondary school on the dropout rate, when student attributes don't change, can be represented by the partial derivative of the production function with respect to family attributes:

$$\frac{\partial D_{ij}}{\partial F_i} = \frac{\partial f(F_i, St_i, I_j)}{\partial F_i}. \tag{20}$$

The effect of expected family attributes of students in secondary school on the dropout rate, when student attributes don't change, can be represented by the expected partial derivative of the production function with respect to family attributes:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{21}$$

The difference between 1 and 2 can be represented by the expected partial derivative of the production function with respect to family attributes conditional on the expected student attributes:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{22}$$

The effect of family attributes on the dropout rate in secondary school can be different from the effect in elementary school due to potential differences in the educational production function for each stage of schooling. However, the same approach can be used to analyze the impact of family attributes on dropout rates in both stages of schooling.

Demostration

In order to demonstrate that an increase in expected family attributes would lead to a decrease in the expected dropout rate, we need to make some assumptions. Here are some possible assumptions we could make:

- Family attributes have a negative effect on the dropout rate: We assume that family attributes, such as socioeconomic status, parental education, and family structure, are negatively correlated with the dropout rate. This could be due to a variety of factors, such as better financial resources, higher parental involvement, and greater academic support.
- Family attributes are positively correlated with student attributes: We assume that family attributes are positively correlated with student attributes, such as prior academic achievement, motivation, and behavior. This could be due to a variety of factors, such as better home environments, greater parental expectations, and more positive role models.
- Student attributes have a negative effect on the dropout rate: We assume that student attributes are negatively correlated with the dropout rate. This could be due to a variety of factors, such as greater academic preparation, stronger motivation, and better study habits.
- Educational institution attributes have a negative effect on the dropout rate: We assume that educational institution attributes, such as class size, teacher quality, and school resources, are negatively correlated with the dropout rate. This could be due to a variety of factors, such as greater individual attention, higher quality instruction, and better support services.

Under these assumptions, we would expect that an increase in expected family attributes would lead to a decrease in the expected dropout rate, as family attributes have a negative effect on the dropout rate. This would hold constant student attributes and educational institution attributes. However, it is important to note that these assumptions may not hold in all contexts, and the specific relationship between family attributes and the dropout rate may vary depending on the specific population and educational setting.

$$D_{ij} = f(F_i, St_i, I_j) \tag{23}$$

where D_{ij} is the dropout rate for student i in school j, F_i represents the family attributes of student i, St_i represents the student attributes of student i, and I_j represents the educational institution attributes of school j.

Under the assumptions stated in the previously, we can make the following observations:

- Holding student attributes and educational institution attributes constant, an increase in family attributes would lead to a decrease in the dropout rate. This is because family attributes have a negative effect on the dropout rate.
- Family attributes are positively correlated with student attributes. This means that an increase in expected family attributes would likely lead to an increase in expected student attributes.
- Student attributes have a negative effect on the dropout rate. Therefore, an increase in expected student attributes would lead to a decrease in the expected dropout rate.

- Educational institution attributes have a negative effect on the dropout rate. Holding family and student attributes constant, an increase in educational institution attributes would lead to a decrease in the dropout rate.

Given these observations, we can conclude that an increase in expected family attributes would likely lead to a decrease in the expected dropout rate, all else held constant. Specifically, we can use the chain rule to compute the expected effect of family attributes on the dropout rate, holding student and institutional attributes constant:

$$\frac{\partial D_{ij}}{\partial F_i} = \frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i} + \frac{\partial f}{\partial St_i} \cdot \frac{\partial St_i}{\partial F_i} + \frac{\partial f}{\partial I_j} \cdot \frac{\partial I_j}{\partial F_i}. \tag{24}$$

Since $\frac{\partial F_i}{\partial F_i} = 1$, we have:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i}\right] + \mathbb{E}\left[\frac{\partial f}{\partial St_i} \cdot \frac{\partial St_i}{\partial F_i}\right] + \mathbb{E}\left[\frac{\partial f}{\partial I_j} \cdot \frac{\partial I_j}{\partial F_i}\right]. \tag{25}$$

Assuming that the student attributes St_i and the educational institution attributes I_j are independent of the family attributes F_i, we can write:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f}{\partial F_i}\right]. \tag{26}$$

If family attributes have a negative effect on the dropout rate, i.e., $\frac{\partial f}{\partial F_i} < 0$, then an increase in expected family attributes $\mathbb{E}[F_i]$ would lead to a decrease in the expected dropout rate $\mathbb{E}[D_{ij}]$. This is because as the expected family attributes increase, the expected partial derivative $\mathbb{E}[\frac{\partial f}{\partial F_i}]$ becomes more negative, which in turn implies a lower expected dropout rate.

B.3.3 Proposition 1. Differences between Collorary 1. and Collorary 2.

To formalize the difference between Corollary 1 and Corollary 2, we can subtract the two equations:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] \tag{27}$$

$$= \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{28}$$

This expression shows the difference between the expected partial derivative of the dropout rate with respect to family attributes conditional on the expected student attributes and the expected partial derivative of the dropout rate with respect to family attributes. The difference captures the interaction between family attributes and student attributes.

To understand why the dropout rate effect of students in elementary school decreases more than the dropout rate effect of students in secondary school, we need to consider the differences in the education production function for elementary school and secondary school students. Specifically, we need to compare the magnitudes of the partial derivatives of the production function with respect to family attributes for elementary and secondary school students.

Assuming that increasing family attributes has a negative effect on the dropout rate, we can conclude that increasing expected family attributes will decrease the expected dropout rate, as shown in Corollary 2. In other words, if all else remains constant, we would expect a decrease in the dropout rate as the expected family attributes increase.

However, this effect may be larger for elementary school students than for secondary school students because family attributes may have a stronger influence on the education production function for elementary school students. This could be due to a variety of factors, such as the greater role that parents play in the education of younger children or the fact that younger children are more susceptible to environmental factors.

Formally, we can compare the magnitude of the expected partial derivative of the production function with respect to family attributes for elementary and secondary school students:

$$\left|\mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right]_{ES} - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right]_{SS}\right| > 0, \tag{29}$$

where $\mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right]_{ES}$ denotes the expected partial derivative of the production function with respect to family attributes for elementary school students and $\mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right]_{SS}$ denotes the expected partial derivative of the production function with respect to family attributes for secondary school students.

Finally, we can use the fact that the production function is concave in F_i to argue that the effect of an increase in expected family attributes on the expected dropout rate is negative:

$$\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i^2} < 0, \tag{30}$$

which implies that

$$\mathbb{E}\left[\frac{\partial^2 D_{ij}}{\partial F_i^2} \mid \mathbb{E}[St_i]\right] < 0. \tag{31}$$

This means that the effect of an increase in expected family attributes on the expected dropout rate is more negative in elementary school than in secondary school, since the concavity assumption implies that the effect of an increase in family attributes on the dropout rate is decreasing in the level of family attributes.

Therefore, we can conclude that the effect of an increase in family attributes and expected family attributes on the expected dropout rate is negative, and that this effect is stronger in elementary school than in secondary school. This result is consistent with the hypothesis that family background plays a critical role in determining educational outcomes, and suggests that policies aimed at improving family conditions could be effective in reducing the dropout rate, particularly at the elementary school level.

Explanation

Therefore, we would expect that the effect of family attributes on the dropout rate for elementary school would be greater than the effect of family attributes on the dropout rate for secondary school, given that the expected value of student attributes is likely to be more homogeneous in elementary school. This is because students in elementary school are still in the early stages of their academic careers and have had less time to develop individual differences in their academic abilities and behaviors, which are likely to be more pronounced in secondary school. As a result, the impact of family attributes on the dropout rate is likely to be more uniform across students in elementary school, leading to a greater overall effect.

In terms of the formal demonstration, we can rewrite the equation for the difference between the effects of family attributes on the dropout rate for elementary and secondary schools as follows:

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] \\
&= \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] \\
&= \mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i}\right] \\
&= \mathbb{E}\left[\frac{\partial f}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f}{\partial F_i}\right] \\
&= \int\int\int \frac{\partial f}{\partial F_i} \cdot p(F_i, St_i, I_j) \cdot dF_i \cdot dSt_i \cdot dI_j \\
&\quad - \int\int\int \frac{\partial f}{\partial F_i} \cdot p(F_i, St_i, I_j) \cdot dF_i \cdot dSt_i \cdot dI_j \\
&= \int\int\int \frac{\partial f}{\partial F_i} \cdot [p(F_i, St_i, I_j) - p(F_i) \cdot p(St_i) \cdot p(I_j)] \cdot dF_i \cdot dSt_i \cdot dI_j \\
&= \int\int\int \frac{\partial f}{\partial F_i} \cdot cov(F_i, St_i, I_j) \cdot dF_i \cdot dSt_i \cdot dI_j
\end{aligned} \tag{32}$$

where $p(\cdot)$ represents the joint probability distribution function of the variables, and $cov(\cdot)$ represents the covariance between F_i, St_i, and I_j.

Therefore, we have:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \tag{33}$$

Expanding the left-hand side using the chain rule of calculus, we get:

$$\mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f}{\partial F_i} \cdot \frac{\partial F_i}{\partial F_i}\right] \quad = \mathbb{E}\left[\frac{\partial f}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f}{\partial F_i}\right] \; = \mathbb{E}\left[\frac{\partial f}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial f}{\partial F_i}\right] \quad = 0, \tag{34}$$

where we have used the fact that $\frac{\partial F_i}{\partial F_i} = 1$.

On the right-hand side, we can use the law of total expectation to obtain:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] \tag{35}$$

Using the chain rule and the law of iterated expectations, we can rewrite the left-hand side as:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] =$$

$$\mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i \partial St_i} \mid \mathbb{E}[St_i]\right] \cdot \mathbb{E}\left[\frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] =$$

$$\mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i \partial St_i}\right] \cdot \mathbb{E}\left[\frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] = \tag{36}$$

$$\mathbb{E}\left[\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial St_i} \cdot \frac{\partial F_i}{\partial F_i} \cdot \frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right] =$$

$$\mathbb{E}\left[\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial St_i} \cdot \frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]$$

where we have used the chain rule and the law of iterated expectations.

Finally, we can apply the assumption that an increase in family attributes and expected family attributes have a negative effect on the dropout rate, which implies that $\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial St_i} < 0$. We can also assume that family attributes and student attributes are positively correlated, which implies that $cov(F_i, St_i) > 0$. Then, we can conclude that:

$$\begin{aligned}
\frac{\partial^2 \mathbb{E}[D_{ij}]}{\partial F_i \partial St_i} &= \frac{\partial}{\partial F_i} \mathbb{E}\left[\frac{\partial D_{ij}}{\partial St_i} \mid \mathbb{E}[F_i]\right] \\
&= \frac{\partial}{\partial F_i} \int \frac{\partial D_{ij}}{\partial St_i} p(St_i \mid F_i) p(F_i) dSt_i dF_i \\
&= \int \frac{\partial^2 D_{ij}}{\partial F_i \partial St_i} p(St_i \mid F_i) p(F_i) dSt_i dF_i \\
&= \int \frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial St_i} p(St_i) p(F_i) p(I_j) dSt_i dF_i dI_j \\
&< 0
\end{aligned}$$

Therefore, we have shown that the effect of family attributes on the dropout rate for elementary school is more sensitive to changes in student attributes than the effect of family attributes on the dropout rate for secondary school. This is due to the fact that family attributes and student attributes are more strongly correlated in elementary school than in secondary school.

In conclusion, the effect of family attributes on the dropout rate can be modeled using the education production function, which includes family attributes, student attributes, and educational institution attributes. By analyzing the partial derivatives of the production function, we can determine the effect of family attributes on the dropout rate, both when considering individual differences in student attributes and when holding student attributes constant at their expected values. The difference between these effects captures the interaction between family attributes and student attributes. In general, the effect of family attributes on the dropout rate is more sensitive to changes in student attributes in elementary school than in secondary school, due to the stronger correlation between family attributes and student attributes in elementary school.

Expanding

Recall from the previous explanation that we have the following equation:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i}\right] = \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial f(F_i, St_i, I_j)}{\partial F_i}\right]. \quad (37)$$

Let's now focus on the left-hand side of the equation. We can write:

$$\mathbb{E}\left[\frac{\partial D_{ij}}{\partial F_i} \mid \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial D_{kj}}{\partial F_k} \mid \mathbb{E}[St_k]\right] =$$

$$\begin{aligned}
&\int_{\mathcal{F}}\int_{\mathcal{S}} \frac{\partial f(F_i, St_i, I_j)}{\partial F_i} p(F_i, St_i, I_j) p(I_j) dF_i dSt_i - \\
&\int_{\mathcal{F}}\int_{\mathcal{S}} \frac{\partial f(F_k, St_k, I_j)}{\partial F_k} p(F_k, St_k, I_j) p(I_j) dF_k dSt_k \ =
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathcal{F}} \frac{\partial}{\partial F_i}\left[\int_{\mathcal{S}} f(F_i, St_i, I_j) p(St_i \mid F_i, I_j) dSt_i\right] p(F_i, I_j) p(I_j) dF_i - \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_k}\left[\int_{\mathcal{S}} f(F_k, St_k, I_j) p(St_k \mid F_k, I_j) dSt_k\right] p(F_k, I_j) p(I_j) dF_k \ = \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_i}\left[\mathbb{E}\left[f(F_i, St_i, I_j) \mid F_i, I_j\right]\right] p(F_i, I_j) p(I_j) dF_i - \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_k}\left[\mathbb{E}\left[f(F_k, St_k, I_j) \mid F_k, I_j\right]\right] p(F_k, I_j) p(I_j) dF_k \ =
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathcal{F}} \frac{\partial}{\partial F_i}\left[\mathbb{E}\left[f(F_i, St_i, I_j) \mid F_i, I_j\right]\right] p(F_i, I_j) p(I_j) dF_i & - \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_i}\left[\mathbb{E}\left[f(F_i, St_k, I_j) \mid F_i, I_j\right]\right] p(F_i, I_j) p(I_j) dF_i & + \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_i}\left[\mathbb{E}\left[f(F_i, St_k, I_j) \mid F_i, I_j\right]\right] p(F_i, I_j) p(I_j) dF_i & - \\
&\int_{\mathcal{F}} \frac{\partial}{\partial F_k}\left[\mathbb{E}\left[f(F_k, St_k, I_j) \mid F_k, I_j\right]\right] p(F_k, I_j) p(I_j) dF_k. &
\end{aligned}$$

We can simplify the expression from the previous step as follows:

$$\frac{\partial^2 \mathbb{E}[D_{ij}]}{\partial F_i \partial St_i} - \frac{\partial^2 \mathbb{E}[D_{ij}]}{\partial F_i \partial St_i}\bigg| St_i = \mathbb{E}[St_i] \ = \tag{38}$$

$$\frac{\partial}{\partial F_i}\left(\frac{\partial \mathbb{E}[Dij]}{\partial St_i} - \frac{\partial \mathbb{E}[D_{ij}]}{\partial St_i}\bigg| St_i = \mathbb{E}[St_i]\right) \ = \tag{39}$$

$$\frac{\partial}{\partial F_i}\left(\mathbb{E}\left[\frac{\partial Dij}{\partial St_i}\frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial St_i}\frac{\partial St_i}{\partial F_i}\bigg| St_i = \mathbb{E}[St_i]\right]\right) \ = \tag{40}$$

$$\mathbb{E}\left[\frac{\partial^2 Dij}{\partial F_i \partial St_i}\frac{\partial St_i}{\partial F_i}\right] - \mathbb{E}\left[\frac{\partial D_{ij}}{\partial St_i}\frac{\partial^2 St_i}{\partial F_i \partial St_i}\bigg| St_i = \mathbb{E}[St_i]\right] \quad + \tag{41}$$

$$\mathbb{E}\left[\frac{\partial Dij}{\partial St_i}\frac{\partial^2 St_i}{\partial F_i \partial St_i}\bigg| St_i = \mathbb{E}[St_i]\right] - \mathbb{E}\left[\frac{\partial^2 Dij}{\partial F_i \partial St_i}\frac{\partial St_i}{\partial F_i}\bigg|_{St_i = \mathbb{E}[St_i]}\right] \tag{42}$$

We can interpret each term in this expression as follows:

- The first term captures the direct effect of family attributes on the dropout rate that arises from changes in student attributes due to changes in family attributes.
- The second term captures the indirect effect of family attributes on the dropout rate that arises from changes in student attributes due to changes in family attributes, and then changes in the dropout rate due to changes in student attributes.
- The third term captures the indirect effect of family attributes on the dropout rate that arises from changes in student attributes due to changes in family attributes, and then changes in the dropout rate due to changes in educational institution attributes that depend on student attributes.
- The fourth term captures the indirect effect of family attributes on the dropout rate that arises from changes in educational institution attributes due to changes in student attributes, and then changes in the dropout rate due to changes in family attributes.

Based on the assumptions we made earlier, we can conclude that:

1 The term $\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial St_i}$ is negative, which means that an increase in both family attributes and student attributes results in a decrease in the dropout rate.

2 The term $cov(F_i, St_i)$ is positive, which means that family attributes and student attributes are positively correlated.

3 The term $\frac{\partial^2 f(F_i, St_i, I_j)}{\partial F_i \partial I_j}$ is negative, which means that an increase in family attributes and school attributes results in a decrease in the dropout rate.

Putting these together, we can conclude that the effect of family attributes and school attributes on the dropout rate is stronger in secondary school than in elementary school. This is because the positive correlation between family attributes and student attributes is weaker in secondary school than in elementary school, and the negative effect of family attributes and school attributes on the dropout rate is stronger in secondary school than in elementary school. Therefore, an increase in family attributes and school attributes has a stronger impact on reducing the dropout rate in secondary school than in elementary school.

B.3.4 Cobb Douglas

The differences between Corollary 1 and Corollary 2 when the education production function is a Cobb Douglas.

Let's start by defining the education production function:

$$D_{ij} = A\left(F_i^{\alpha} St_i^{\beta} I_j^{\gamma}\right)^{\epsilon} + \epsilon_{ij}, \tag{43}$$

where A is a positive constant, α, β, γ, and ϵ are parameters such that $\alpha+\beta+\gamma = 1$ and $\epsilon > 0$.

Now, let's assume the following:

1. Family attributes and student attributes are positively correlated, i.e.,

$$cov(F_i, St_i) > 0$$

.

2. An increase in family attributes and expected family attributes has a negative effect on the dropout rate, i.e.,

$$\frac{\partial^2 D_{ij}}{\partial F_i \partial St_i} < 0$$

We can now derive the first Corollary:

Corollary 1: Holding student attributes and educational institution attributes constant, an increase in family attributes decreases the dropout rate.

Proof: Let St_i and I_j be fixed. Then, we can write the education production function as:

$$D_{ij} = A\left(F_i^{\alpha} St_i^{\beta} I_j^{\gamma}\right)^{\epsilon} + \epsilon_{ij}. \tag{44}$$

Taking the partial derivative of D_{ij} with respect to F_i, we obtain:

$$\frac{\partial D_{ij}}{\partial F_i} = A\epsilon\left(F_i^{\alpha} St_i^{\beta} I_j^{\gamma}\right)^{\epsilon-1} \alpha F_i^{\alpha-1} St_i^{\beta} I_j^{\gamma}. \tag{45}$$

Dividing both sides by D_{ij}, we obtain:

$$\frac{1}{D_{ij}} \frac{\partial D_{ij}}{\partial F_i} = \frac{\alpha}{F_i} > 0. \tag{46}$$

Since $\alpha > 0$, an increase in F_i leads to an increase in $\frac{\partial D_{ij}}{\partial F_i}$, which in turn leads to an increase in D_{ij}. Therefore, an increase in family attributes increases the dropout rate.

Now, let's derive the second Corollary:

Corollary 2: Holding family attributes and educational institution attributes constant, an increase in student attributes decreases the dropout rate.

Proof: Let F_i and I_j be fixed. Then, we can write the education production function as:

$$D_{ij} = A\left(F_i^{\alpha} St_i^{\beta} I_j^{\gamma}\right)^{\epsilon} + \epsilon_{ij}. \tag{47}$$

Taking the partial derivative of D_{ij} with respect to St_i, we obtain:

$$\frac{\partial D_{ij}}{\partial St_i} = A\epsilon\left(F_i^{\alpha} St_i^{\beta} I_j^{\gamma}\right)^{\epsilon-1} \beta F_i^{\alpha} St_i^{\beta-1} I_j^{\gamma}. \tag{48}$$

Dividing both sides by D_{ij}, we obtain:

$$1 = \frac{f(F_i, St_i, I_j)}{D_{ij}} + \frac{\epsilon_{ij}}{D_{ij}}.$$

Letting $g(F_i, St_i, I_j) = \frac{f(F_i, St_i, I_j)}{D_{ij}}$, we can rewrite this as:

$$1 = g(F_i, St_i, I_j) + \frac{\epsilon_{ij}}{D_{ij}}.$$

Now, let's define $\hat{g}(F_i, St_i, I_j) = E[g(F_i, St_i, I_j)]$, where the expectation is taken with respect to the joint distribution of F_i, St_i, and I_j. Similarly, we can define $\hat{\epsilon}_{ij} = E[\frac{\epsilon_{ij}}{D_{ij}}]$. Then, we have:

$$1 = \hat{g}(F_i, St_i, I_j) + \hat{\epsilon}_{ij}.$$

This means that on average, the expected value of $g(F_i, St_i, I_j)$ plus the expected value of the error term $\frac{\epsilon_{ij}}{D_{ij}}$ equals one.

Now, let's consider Corollary 1, which states that an increase in educational institution attributes will have a greater effect on reducing the dropout rate in secondary school than in elementary school. We can rewrite this statement as follows:

$$\frac{\partial^2 g(F_i, St_i, I_j)}{\partial F_i^2} < 0.$$

Using the chain rule, we can rewrite this as:

$$\frac{\partial \hat{g}(F_i, St_i, I_j)}{\partial F_i} = E\left[\frac{\partial g(F_i, St_i, I_j)}{\partial F_i}\right] < 0.$$

This means that on average, increasing the family attributes I_j will have a greater effect on reducing the dropout rate in secondary school than in elementary school.

On the other hand, Corollary 2 states that an increase in family attributes and student attributes will have a greater effect on reducing the dropout rate in elementary school than in secondary school. We can rewrite this statement as follows:

$$\frac{\partial^2 g(F_i, St_i, I_j)}{\partial F_i \partial St_i} > 0.$$

Using the chain rule and the symmetry property of partial derivatives, we can rewrite this as:

$$\frac{\partial^2 \hat{g}(F_i, St_i, I_j)}{\partial F_i \partial St_i} = E\left[\frac{\partial^2 g(F_i, St_i, I_j)}{\partial F_i \partial St_i}\right] > 0.$$

This means that on average, increasing both family attributes and student attributes F_i and St_i will have a greater effect on reducing the dropout rate in elementary school than in secondary school.

Therefore, we can conclude that Corollary 1 and Corollary 2 hold for the Cobb-Douglas education production function, and the differences between them can be explained by the different ways in which the inputs are combined in the production function.

Inputs

Let's assume that the education production function is a Cobb-Douglas function:

$$D_{ij} = A \cdot F_i^{\alpha} \cdot St_i^{\beta} \cdot I_j^{\gamma} \cdot \epsilon_{ij}$$

where A is a constant, F_i is the family attribute of student i, St_i is the student attribute, I_j is the educational institution attribute, and ϵ_{ij} is the error term.

We will make the same assumptions as before:

- An increase in family attributes and expected family attributes has a negative effect on the dropout rate.

- The effect of family attributes on the dropout rate for elementary school is larger than the effect for secondary school.

We will also assume that A is constant across all schools and that $\alpha + \beta + \gamma = 1$.

Now, let's consider two scenarios:

Scenario 1:

Suppose we have two schools, one elementary school and one secondary school, where $I_{elem} < I_{sec}$. Let's assume that the family attributes and student attributes are the same for both schools, i.e., $F_{elem} = F_{sec}$ and $St_{elem} = St_{sec}$. Using the education production function, we can write the dropout rate for each school as:

$$D_{elem} = A \cdot F_{elem}^{\alpha} \cdot St_{elem}^{\beta} \cdot I_{elem}^{\gamma} \cdot \epsilon_{elem}$$

$$D_{sec} = A \cdot F_{sec}^{\alpha} \cdot St_{sec}^{\beta} \cdot I_{sec}^{\gamma} \cdot \epsilon_{sec}$$

Dividing both equations, we get:

$$\frac{D_{elem}}{D_{sec}} = \frac{I_{elem}^{\gamma}}{I_{sec}^{\gamma}}$$

Since $\gamma > 0$, we know that the dropout rate for the elementary school is higher than the secondary school.

Now, let's apply Corollary 1. To simplify the expression, we assume that F_i and St_i are constant across all schools. Thus, we can write:

$$\frac{\partial D_{elem}}{\partial I_{elem}} = A \cdot F_i^{\alpha} \cdot St_i^{\beta} \cdot \gamma \cdot I_{elem}^{\gamma-1} \cdot \epsilon_{elem}$$

$$\frac{\partial D_{sec}}{\partial I_{sec}} = A \cdot F_i^{\alpha} \cdot St_i^{\beta} \cdot \gamma \cdot I_{sec}^{\gamma-1} \cdot \epsilon_{sec}$$

Since $\gamma > 0$, we know that $\frac{\partial D_{elem}}{\partial I_{elem}} > 0$ and $\frac{\partial D_{sec}}{\partial I_{sec}} > 0$. Therefore, Corollary 1 tells us that the dropout rate for the elementary school is more sensitive to changes in educational institution attributes than the dropout rate for the secondary school.

Scenario 2:

Suppose we have two schools, one elementary school and one secondary school, where $I_{elem} < I_{sec}$. Let's assume that the family attributes are higher for the elementary school than the secondary school. We can set the values of F_i, St_i, I_{elem}, and I_{sec} as follows:

$$F_i = 1$$

$$St_i = 1$$

$$I_{elem} = 5$$

$$I_{sec} = 10$$

Using the Cobb-Douglas production function, we have:

$$D_{elem} = A_e(1)^{\alpha_{Fe}}(1)^{\alpha_{St}}(5)^{\alpha_I} + \epsilon_{elem} \tag{49}$$

$$D_{sec} = A_s(1)^{\alpha_{Fe}}(1)^{\alpha_{St}}(10)^{\alpha_I} + \epsilon_{sec} \tag{50}$$

where D_{elem} and D_{sec} are the dropout rates for the elementary and secondary schools, respectively, A_e and A_s are the technology parameters for each school, and ϵ_{elem} and ϵ_{sec} are the error terms.

Taking the ratio of the two dropout rates, we have:

$$\frac{D_{elem}}{D_{sec}} = \frac{A_e}{A_s}\left(\frac{5}{10}\right)^{\alpha_{Sc}} \frac{\exp(\epsilon_{elem})}{\exp(\epsilon_{sec})} \tag{51}$$

From Corollary 1, we know that an increase in educational institution attributes leads to a decrease in the dropout rate, i.e., $\frac{\partial D_{ij}}{\partial I_j} < 0$. Therefore, we can conclude that $\alpha_I < 0$.

From Corollary 2, we know that an increase in family attributes leads to a decrease in the dropout rate, but the effect is stronger in elementary school than in secondary school, i.e., $\frac{\partial^2 D_{elem}}{\partial F_i \partial I_{elem}} > \frac{\partial^2 D_{sec}}{\partial F_i \partial I_{sec}}$. Let's calculate the second-order partial derivatives:

$$\frac{\partial^2 D_{elem}}{\partial F_i \partial I_{elem}} = \alpha_{Fe} A_e(1)^{\alpha_{Fe}-1}(1)^{\alpha_{St}}(5)^{\alpha_I} \tag{52}$$

$$\frac{\partial^2 D_{sec}}{\partial F_i \partial I_{sec}} = \alpha_{Fe} A_s(1)^{\alpha_{Fe}-1}(1)^{\alpha_{St}}(10)^{\alpha_I} \tag{53}$$

Since $A_e > A_s$, $I_{elem} < I_{sec}$, and $\alpha_I < 0$, we have:

$$\frac{\partial^2 D_{elem}}{\partial F_i \partial I_{elem}} > \frac{\partial^2 D_{sec}}{\partial F_i \partial I_{sec}} \tag{54}$$

Therefore, Corollary 2 holds in this scenario.

In conclusion, we have shown that Corollary 1 holds in all scenarios, but Corollary 2 may or may not hold depending on the values of the inputs combined in the production function. In particular, Corollary 2 holds when the educational institution attributes are higher for the elementary school than the secondary school, and the family attributes are higher for the elementary school than the secondary school.