Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 256

Sweet child o` mine: The Impact of mining on educational and labor market outputs in Colombia

Luis Omar Herrera Prada

February 2024

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany · Phone: +49-(0)551-3928172 · Fax: +49-(0)551-3928173

e-mail: uwia@gwdg.de · http://www.iai.wiwi.uni-goettingen.de

Sweet child o' mine:

The Impact of mining on educational and labor market outputs in Colombia

Luis Omar Herrera Prada[a]*

[a]*IAI, Universität Göttingen, Göttingen, Germany*

International Monetary Fund, Washington DC, USA

ORCiD 0000-0003-1099-8102

*Corresponding Author: l.herreraprada@stud.uni-goettingen.de Twitter: luomar

Address: 700 19th ST NW HQ2-11 Washington DC 20431

Sweet child o' mine:
The Impact of mining on educational and labor market outputs in Colombia

This paper examines the effects of mineral extraction on human capital formation in Colombia, a country rich in natural resources but struggling with low college attendance, high youth unemployment, and high informality in the labor market. Leveraging the allocation of natural resources as a quasi-experimental setting, we link administrative data for 5.5 million secondary school graduates from 2002 to 2014 with information on legal mines in 2014 based on the distances from their respective schools. Employing Instrumental Variables and a Differences-in-Differences approach, we identify treated individuals as graduates from secondary schools located closest to operational mines at the time of their graduation. The findings indicate that active mines positively influence school cohort sizes, student academic performance, and enrollment in higher education. However, they also negatively impact entry into the formal labor market, particularly for roles associated with extractive industries. Substantial heterogeneity exists in the outcomes associated with the various extracted products, leading to the identification of distinct categories: "good mines" and "bad mines."

Keywords: Mining, Labor Market, Education

JEL Classification Numbers: I23 I24 J24 L71 L72

Introduction

This paper contributes to the literature by separating the regional effects within political boundaries from the effects caused by the mine and extending results to other materials. So far, the literature has made progress in measuring the impacts of mines on academic performance and other labor market outputs but has not been able to separate the effect of the region. It also complements Bonilla Mejía (2020) on exploiting the labor market mechanisms and the distances between mines and schools. Since mines leave royalties and generate policies evidenced in the academic results, the transmission mechanism is mainly the region and not necessarily by the exposure of the school to mine in operation. we address this by controlling for political boundaries and the influence of the mine on the school, depending on the distance between the school and the mine and the number of mines around the school. In this way, we allow a mine located in a specific region to have effects on a school located in another, separating the regional effects from the direct influence of the mine on the student's academic results.

Latin America and the Caribbean (LAC) is a world region with abundant natural resources and complex institutional, political, financial, economic, and social characteristics that have impacted its economic development. In Colombia, the exploitation of oil and minerals is a significant part of the main economic activities: more than 33.2% of its exports are from oil and 21.6% from minerals, according to the National Department of Statistics (DANE).

The literature regarding the effect of natural resource exploitation from mining activities on economic development indicators is broad. It explains why governments concentrate their investments in this activity instead of promoting other sectors Gylfason (2001). Several papers have suggested strategies or found interesting results on the consequences of exploiting natural resources for social dynamics (Bonilla Mejía, 2020; van der Ploeg, 2011), economic growth (Barro, 2001;

Martínez Ortiz & Aguilar, 2012), poverty (Litschig & Morrison, 2013; Loayza & Rigolini, 2016; Pegg, 2006), government efficiency in the provision of public goods (Angrist & Kugler, 2008; Caselli & Michaels, 2013; Loayza et al., 2014; L. R. Martínez, 2023), and quality of education (Agüero et al., 2016; Álvarez & Vergara, 2022; Hanushek & Woessmann, 2010). However, it is essential to discuss the relevance of the geographical location where the activity is carried out and the impact generated by its exposure to the people (Torvik, 2002). In the literature, the results regarding educational outcomes are more related to pollution-related school absences due to cognitive skills impairments during "the school stage" (Almond et al., 2009; Currie et al., 2009; Park et al., 2002).

Using administrative data for Colombia, Instrumental Variables (IV) and Differences in Differences approaches, this paper provides evidence of the link between mining and some educational and labor market outputs in Colombia. we exploit the mines' locations as a natural experiment and the operation of the closest mine to each school in Colombia as the treatment for the nation's secondary graduates from 2002 to 2012. we use the universe of secondary school graduates in Colombia from 2002 to 2012 merged with information from the Ministry of Education and Ministry of Health to mark those who went to college and/or the formal labor market, respectively. All legal mines and schools were geo-localized, creating a matrix of distances to assign the closest mine's information to each school. The students who graduated from a secondary school whose nearest mine was operating during their graduation year are the treated group. Assigning the treatment by using the closest mine to each school allows to avoid the correlation between the municipalities' information and their outcomes. It is vital to separate the effects of the mines and the regional budget; the closest mine can be crossing the border of the country's domestic political division.

In line with Angrist & Kugler (2008) results, we find that the size of cohorts increases around 6% (LATE) and 7.2% (ATT) when the nearest mine is in operation. The Saber 11 test score increases if the nearest mine is in operation by 3.87 points (an improvement of 8.2% compared to the control mean), reaching up to 17 points 9 years after the opening of the mine. we also found positive results for the probability of enrolling in higher education; if the nearest mine is in operation, the likelihood of enrolling in higher education increases by 4.5% (LATE) and 12.2% (ATT). These results are impressive because the same effect that causes an increase in the size of the cohorts does not affect the quality of the education since the Saber 11 score does not decrease. It also boosts the probability of pursuing higher education. Finally, the results show that if the mine closest to the schools is in operation, the probability of having a formal job drops between 0.2% and 8.6%.

The rest of the document is organized as follows: Section 2 presents a literature review, Section 3 describes the data, Section 4 outlines the modeling strategy, Section 5 reports the results, and Section 6 offers the conclusions.

Literature Review

In this section, we describe the literature related to the chapter. First, we discuss how the broad literature analyzes and links different macroeconomic topics, natural resources, and educational outcomes. Next, we present some specific papers related to the Colombian case and discuss the current context.

Since mining is one of the largest contributors to State revenues, and public policy tends to be in line with the incentives for foreign direct investment (Gylfason, 2001), mining regulation should be a priority in the government's agenda. With effective regulation, governments can also control the collateral side effects of mining activity. However, mining activities' effects in LAC are dual: although there is evidence of an improvement in macroeconomic indicators (Loayza et al., 2014; Maldonado, 2018), the mining industry has increased social inequality.

On the one hand, mining activity encourages per capita consumption and employment, with a significant spillover in education. The mine's labor force will require better-educated immigrants such as operators, technicians, engineers, ecologists, who would move to the area and will require improvements in the region's education and health system quality for their families (Loayza & Rigolini, 2016).

On the other hand, the returns generated by natural resources are about 60% of local governments' revenue in developing countries. However, the perception is that their distribution is inefficient in providing public services such as health and education. Regarding education, a deterioration of these services can help explain the decrease in cognitive abilities and increase in school absenteeism (Gilliland et al., 2001; Lavy & Roth, 2014); regarding health, the community may be affected by exposure to pollution generated by extractive activity (L. R. Martínez, 2023; Z. Martínez et al., 2017; Romero & Saavedra, 2016).

The related literature suggests some strategies that explain and interpret in different contexts the consequences of the exploitation of natural resources for social dynamics and economic growth (Angrist & Kugler, 2008; Loayza et al., 2014; van der Ploeg, 2011). For instance, different results or behaviors have been identified in terms of economic policy (Caselli, 2006), including modest reductions in poverty coupled with a positive impact on literacy rates (Litschig & Morrison, 2013; Pegg, 2006). Moreover, the government's efficiency in the provision of public goods has been scrutinized (Maldonado, 2018; Maldonado & Ardanaz, 2023). Along these lines, it is worth highlighting Z. Martínez et al. (2017), who compares the effects of public spending from two types of returns–oil revenues (transfers) and the fiscal effort of local governments–and finds equivalences in the spending on education as a result of both sources of income.

In this context, Agüero et al. (2016) find that the distribution of transfers from mining activity in a "boom" context directly impacts the conditions of public services provided to the population, especially in the quality of the education, observed via mathematics' scores, as Hanushek & Woessmann (2010) predicts. Also, Álvarez & Vergara (2022) complement these findings by introducing the change in wages in the producing regions as an important transmission mechanism for human capital formation, measured in years of schooling, and the high demand for work during the "boom" of natural resources exports. According to economic growth theory, there are two ways in which human capital impacts growth: through the number of years of education (years of schooling), and via the effect of the quality of education on economic growth - productivity - (Barro, 2001). On this subject, Angrist & Kugler (2008) find that an increase in the demand for coca, opium, and diamond production increases school attendance and child labor in the same rural areas where such goods are concentrated during "boom" times in Colombia. This result is linked to civil conflicts and the violence associated with growing illegal crops–mainly coca (Currie et al., 2014; Loayza et al., 2014; Ross, 2015). This link to civil conflicts is also one reason that education results are not good, particularly in rural areas (Dube & Vargas, 2013).

However, the impact of mines' locations on human capital formation and entry into the labor market remains ambiguous, with investment changes only partially accounting for the observed outcomes. Geographical location and exposure to mining activities play a crucial role, resulting in both gains and losses for the affected population (Torvik, 2002). Notably, the quality of and access to health systems in these areas undergo significant changes, leading to increased pollution and toxic emissions (Currie et al., 2014; Romero & Saavedra, 2016). These environmental and health concerns contribute to heightened school absences, attributed to compromised health and impaired cognitive skills during the schooling phase (Almond et al., 2009; Currie et al., 2009; Park et al., 2002).

Finally, one key paper is Bonilla Mejía (2020), it explores how gold mining (legal and illegal) impacts human capital accumulation through two main measures: an invariable interaction over time between the intensity of gold deposits near schools and the recording of international gold prices, and two alternatives for assessing the differential effects of illegal mining (active mining titles and mining deforestation). Its objective is to elucidate the labor market, especially violence and corruption, as mechanisms through which mining may affect human capital accumulation (enrollment, dropout rate, and results of standardized tests or household surveys). Using differences in differences and Instrumental Variables models, Bonilla Mejía (2020) finds that mining increases enrollment and progression (promotion rates between levels and reduces dropouts) at lower school levels. These effects tend to fade at the upper secondary level. Illegal mining shows larger but consistent effects on extensive outcomes (multiplied up to 3 times when including instrumental variables) for gold mines located between 10 and 50 km from the school. The effects are higher for the elementary school level located closer than 30 km and for the secondary schools located between 30 and 50 km from the mines. In general, gold mining decreases student performance at school, especially in the early stages; the impacts are three times higher when including (illegal) mining deforestation.

This research complements Bonilla Mejía (2020) on exploiting the labor market mechanisms and the distances between mines and schools. It incorporates the number of mines surrounding the school to test the strength of the effects. It differed in the effects studied on the size of the cohort, the Saber 11 test score, and the probabilities of enrolling in higher education and the labor market. The controls applied to the models are extended to include family, socioeconomic, and institutional characteristics, and other minerals extracted in the mines. It also includes all high schools, their schedules, sectors, and types. The relevant value-added is the measurement of exposure. This measurement is crucial, as in the literature, it is not very easy to separate the effect of the mine (in this case) and the impacts of the municipality's royalties. It includes the effects of temporal variation in the definition of treatment and exposure. Furthermore, doing so, also eliminates biases from external issues such as economic benefits from operating the mine in a specific municipality, since the observed effect of the proximity of the mine on nearby schools regardless of the municipality.

Data

This section describes the chapter 's data and outline our assumptions. The first part describes the databases for individuals, schools, and mines. The second part describes how these databases were combined.

In Colombia, the Ministry of Education (MEN) administers the Saber 11 exam, a prerequisite for higher education enrollment, to all secondary education students through the Colombian Institute for the Promotion of Higher Education (ICFES). The ICFES database (also known as Saber 11 database) contains comprehensive information, including student demographics, exam scores, and various economic, individual, family, and academic variables. While utilizing the Saber 11 database from 2002 to 2016, certain modifications were made. Notably, the absence of household income data for certain periods necessitated imputation based on the mode of household income within the same school for other periods, prioritizing the higher value when multiple options existed. Moreover, the standardized test score scale changed over time, preventing direct comparisons across different years. To address this issue, each student's percentile on the Saber 11 exam was calculated and used as the new standardized score variable. This approach aligns with the Ministry of Education's methodology for standardizing the Saber 11 score in its own information systems. Additionally, key variables, such as the year of birth, gender, standardized Saber 11 test score, household income, school ID code for integration with the school census, and an ethnic group indicator, were extracted from the ICFES database for analysis. The size of cohort was estimated by counting the students that present the test linked to a school code in a period of time (see descriptive stats in Table 1).

The information gathered on secondary schools originates from the Ministry of Education. The dataset encompasses essential details, such as the school's address, geolocation, urban or rural classification, and descriptive information concerning the sector (public or private) and type (academic, technical, or military). Notably, due to varying schedules, multiple schools can share the same campus in Colombia's secondary education system. Consequently, certain schools may share the same geographical location while differing in their respective sectors. For instance, a public-sector building might serve as a secondary school in the morning and be leased to a private school for afternoon or evening sessions. In this scenario, the schools would share the location while differing in their operational sectors. Schools not operated by the government are categorized as private, including those managed by private entities under contract with the government. In terms of the urban/rural classification, the definition of "rural" encompasses any location not explicitly designated as urban. Thus, schools categorized as having mixed urban-rural or rural-urban statuses are coded as rural. From the secondary school census, the following attributes are utilized: school location (as specified above), school sector (as described earlier), school type (e.g., single-gender or coeducational), school shift (i.e., morning, afternoon, or evening classes), and school degree type (academic or technical) (see descriptive stats in Table 1, a map showing the locations of the school is available in Figure 1).

In addition to the ICFES database and the school census, the Colombian Ministry of Education (MEN) manages the System for the Prevention and Analysis of Dropout in Higher Education Institutions (SPADIES) database. This comprehensive database comprises the academic information of all students enrolled in higher education institutions (HEIs) since 1998. SPADIES database includes details such as the HEIs in which students are enrolled, first and last periods of enrollment in higher education, status in the system (active, dropout, graduated), program of study and area of concentration, type of degree (bachelor's or associate programs), and method of learning (classroom or online). This paper uses a merged dataset from the ICFES and SPADIES databases from 2002 to 2017, employing a dummy variable that takes the value of 1 if the student enrolled in higher education (see descriptive stats in Table 1).

The Planilla Integrada de Liquidación de Aportes (PILA) database, overseen by the Ministry of Labor, contains Social Security payment records for all individuals in the formal labor sector. The database includes details such as the number of days worked annually, employment type (e.g., full-time, part-time, self-employed), and employer type (e.g., public company, private company, nonprofit, non-governmental organization, etc.). These data are extracted from the monthly report of contributions made by all formal Colombian workers to pension and health funds (see descriptive stats in Table 1).

This dataset is solely utilized to evaluate the likelihood of an individual entering the formal labor sector subsequent to secondary school. Specifically, to investigate whether proximity to a mine influences the decision to pursue higher education versus entering the formal labor market directly after completing secondary school. A dummy variable is created from the merge between the ICFES database and the PILA database, taking the value of 1 if the student enrolled in the labor market.

For data on all legal mines in Colombia, it uses the Colombian Mining Census, a database collected by *Tierra Minada*, a nonprofit organization that holds the information for the permits and requests to operate mines in Colombia[1]. The Colombian Mining Census contains information such as mine size, natural resource extracted, geographic location, and dates for the start and close of operations. All data for mines that have been in operation at any point between 2002 and 2014 are used.

This database contains 9,545 active mines (valid mining titles) for the period of 2002 to 2014. The database includes an address for each mine, but some of the addresses referred to the offices that managed the mine or to the mining complex's entrance, both of which may have been far from the actual mine. To ensure an accurate location for each mine, we programmed algorithms to analyze Google Earth pixel data to detect each mine's most precise location[2]. The final product was a database with the longitude and latitude for each mine reported in the Census. The descriptive information for this database can be found in Table 1A and a map with their location is available in Figure 2.

Finally, to use in the IV approach, we used the data from the Base Metals Price Index (PMETA) from the (International Monetary Fund, 2023). It includes the prices of Aluminum, Cobalt, Iron Ore, Lead, Molybdenum, Nickel, Tin, Uranium, and Zinc. The IMF estimates the PMETA at least twice per year as part of spring and fall assumptions. The price index's base year is 2016=100 (see descriptive stats in Table 1).

The Administrative Data Matching Process and Final Database

Various merging approaches for the administrative databases were employed, contingent on distinct

---

[1] Data are available at https://sites.google.com/site/tierraminada/

[2] Activisual, a software development company, provided support programming the code, cross-checking on-field some results from the algorithm, and contacting some mines with incomplete contact information.

identification variables. First, we merged the ICFES (Saber 11) and SPADIES databases using the same merging technique that the MEN uses[3]. The merging of Saber 11 and PILA was executed by the Ministry of Health and Social Protection employing the national identification number of Colombia. The merge between Saber 11 and the Census of Schools was achieved using the ICFES' school code. MEN provided the school's location. Activisual got the location of the mine, and it is computed the orthodromic distance between secondary schools and mines in kilometers. Then, the information about the mine closest to each student is assigned to that student's record.

The final database, created at the individual level, originates from the ICFES database, encompassing information on 7,517,983 students who took the Saber 11 exam between 2002 and 2016. To match the Colombian Mining Census data, entries after 2013 were excluded, resulting in 6,172,756 students in the database. Among them, 400,819 students without a linked Secondary School during the Saber 11 test, 60,460 with missing value in the Saber 11 score, and 491 students associated with mines lacking a defined extractable material were dropped. Variables from the ICFES database, including cohort sizes and Saber 11 exam scores, were retained. The SPADIES database was used to obtain higher education enrollment status. Social Security records were used to capture students' labor status in the formal sector after graduating from secondary school.

[3] The algorithm takes two key variables, namely the full name and the date of birth, from the databases. Firstly, the algorithm removes the spaces, converts all alphabetic characters to uppercase, and then decomposes the strings into all possible combinations of the characters. For instance, the name "Tom" is transformed into TOM, MOT, OTM, OMT, TMO, MTO. Next, the algorithm compares each discomposed key variable for every observation in each database to all possible observation matches between the databases. If the comparison reaches a certain "trigger" level, the algorithm identifies the observation as a match. The level of match is the percentage of similarity between the discomposed variables. The algorithm is cautious, meaning that if there is more than one potential matching option, it will not execute the matching. In this paper, the trigger value used is 98%, the same as the value used by the Ministry of Education in the SPADIES-ICFES match.

Additionally, information regarding the nearest mine to each student's record was incorporated, associating students with the mine closest to their respective secondary schools. To ensure result transparency and stability, the approach used the nearest actively operating mine, resolving ambiguities arising from active and non-active mines within specified radius.

The final database comprised 5,710,986 students with data for the variables of interest (size of cohort, score in Saber 11, enrollment in higher education, and enrollment in the labor market). Data also include individual information such as year of birth, gender, family household income, parents' education levels, ethnic minority status, school type, school term duration, school coordinates, and mine size. Non-merge students or those with missing data for the controls or variables of interest are distributed homogeneously across time. Finally, PMETA and other commodities prices were merged by year. The final database is a repeated cross-section at the individual level.

Modeling Strategy

In this section, the model specification is presented. In the first part, we present a ordinary least squares approach (OLS) to have a guide about the sign of de factors that can be affecting any of the five outcomes: (1) size of secondary graduation cohort (2) score on the Saber 11 exam, (3) probability of enrollment in Higher Education, and (4) probability of entry into the formal labor market. In the second part, an Instrumental Variables (IV) approach is developed to obtain causal estimators of the mine's operation on each of the four outcomes mentioned above. Finally, in the third part, we follow the Callaway & Sant'Anna (2021) methodology to aggregate the results of the difference and difference approach.

OLS approach

The model is based on models used by Balza et al. (2021) and Bonilla Mejía (2020). we estimate the impact of mining activity on educational and labor market outcomes using information from the mines that are located near secondary schools. we control by the size of the mine and the extracted product, extending the analysis of Bonilla Mejía (2020) beyond gold mines and complementing the analysis of Balza et al. (2021) in reaching the full extractive sector.

Using school location and detailed information from the Colombian Mining Census, we incorporated the number of mines around each school and the moment when these mines started to operate. In short, the model compares a student's outcomes before and after the operation time of the closest mine to the school. To do so, we created our variable of interest OM = Closest mine is operating as a dummy variable that takes the value of 1 when the closest mine to each school starts its operation OM = 1 and the value of zero OM = 0 when the nearest mine is not operating. Mines can be out of business or not yet working. The treatment is assigned to each student through his or her secondary school and year of graduation; this means that the treatment group consists of those students who took the Saber 11 exam during the mine's period of active operation. The possibility of identifying how one mine simultaneously impacts different schools located in different municipalities helps us avoid selection problems due to the municipality in the results. However, a control by departmental fixed effects is included. Equation 1 shows the regression approach model:

$$Y_{ist} = \alpha_i + \delta_t + \rho OM_{is\,k\leq t} + X'_{ist}\beta + \varepsilon_{ist} \quad (1)$$

In this equation, Y_{ist} represents one of the four outputs that are analyzed: (1) size of secondary graduation cohort (2) score on the Saber 11 exam, (3) probability of enrollment in Higher Education, and (4) probability of entry into the formal labor market. Some tables will show a small variation of output (4) extending it only to those who enrolled in works related to formal labor market in the mining sector. The variable of interest is "OM" a dummy variable that equals 1 if individual "i" graduated from a secondary school "s" whose nearest mine was operating in the year "k" (where k <= t) and 0 otherwise. The parameter ρ denotes the rate of change in output Y_{ist} due to the closest mine's operation. The control α_i incorporates all the time-invariant characteristics of each individual, including gender, Saber 11 score (used as a proxy for academic ability), household income, parents' education, ethnicity, and year of birth. The parameter δ_t captures time-varying drivers. The vector "X" includes observable predictors for the outputs that are linked to the student by his/her school Distance to the closest mine, Distance square, Time of operation of the closest mine, Number of mines in certain ratios, Size of the closest mine, Sector of the school, if school is Coed, if school is in urban area, if school conducts to academic (regular degree), the calendar of the school (starting the academic year in January), the location of the school.

It is important to acknowledge that this model, while aligned with previous research and pertinent to our inquiries, faces limitations in establishing causality. Notably, the non-random assignment of education levels poses a significant challenge in studying the connection between education and earnings. Individuals make deliberate choices concerning their educational paths, considering opportunity costs (as emphasized by Wood (2009)). To mitigate potential econometric challenges such as sample selection and endogeneity, we employ an instrumental variables approach to estimate the Local Average Treatment Effect (LATE) of the mine in operation. Additionally, it is crucial to recognize that effects may vary across different cohorts, as mines commence operations at various locations and times. To address this concern, we utilize Callaway and Sant'Anna (2021)'s framework to estimate the Average Treatment Effect on the Treated (ATT). This strategy enables us to investigate variations in the effects of education on earnings among diverse subgroups.

Instrumental Variables Approach

The use of instrumental variables is an appropriate methodology when addressing potential endogeneity concerns. In this study, we employ a two-stage least squares (2SLS) estimator to systematically tackle these issues. To do so, we utilize the price index for metals from the IMF (PETA) interacted with the number of mines within a 1 km radius of schools in the year preceding the start of their operation plus one; similar to the approach used by Balza et al. (2021), Black et al.

(2005), Bonilla Mejía (2020), Dube & Vargas (2013), and Michaels (2011). This interaction serves as an instrumental variable, representing a proxy for the supply of mines per school, to quantify the probability of the mines commencing operations.

Therefore, our approach involves estimating a first step to predict the probability of start operation, employing our instrument as an independent variable. The vectors of control variables α_i δ_t and X_{ist} remain consistent with Equation 1. The first step equation is formally specified as follows:

$$OM_{ist} = \alpha_i + \delta_t + \mu PETA_t \times (Mines1K_{ist-1} + 1) + X'_{ist}\beta + \varepsilon_{ist} \quad (2)$$

The instrument used in this study is exogenous, as it is derived from a set of prices in the international market, where Colombia has no control over these prices and acts as a price taker. Moreover, mines are not established with direct consideration of schools. Therefore, the instrument can incentivize mine operation (relevance assumption), as an increase in prices would make mining more attractive. Simultaneously, a congested region near schools can discourage mine establishment in certain locations. However, the instrument itself cannot directly impact any of the outputs (exchangeability assumption), as it does not share common causes with the outcomes. International prices and rents do not directly affect students, households, or schools (exclusion restriction).

In Equation 3, we utilize the estimated probability of a mine commencing operations, obtained from Equation 2, as an instrument for "OM". Given that our instrument satisfies the relevance assumption, exchangeability assumption, and exclusion restriction, as explained earlier, the exogenous variation provided by the instrument in the Instrumental Variables (IV) approach yields a precise local average treatment effect (LATE). Hence, the results in Equation 3 can be interpreted as the causal effect of an operational mine on the analyzed output.

$$Y_{ist} = \alpha_i + \delta_t + \rho \widehat{OM}_{is\,k\leq t} + X'_{ist}\beta + \varepsilon_{ist} \quad (3)$$

Heterogeneous Difference in Differences (DiD) Approach

The conventional Difference-in-Differences (DiD) approach typically employs a 2X2 model involving two time periods and two groups. In the initial period (t=0), both groups exhibit similar characteristics and lack exposure to the treatment. In the subsequent period (t=1), some individuals receive the treatment, forming a "treated" group (OM=D=1), while others remain "controls" (OM=D=0) without the treatment. This fundamental model aligns with the interpretation presented in Equation 1, where t=0 corresponds to the year preceding the commencement of the nearest mine's operations, and t=1 represents the subsequent year when the mine begins to operate. Equation 4 outlines the foundational framework for a DiD analysis based on Equation 1.

$$Y_{ist} = \alpha_i + \delta_t + \rho_{is}\mathrm{OM} \times \mathrm{t} + X'_{ist}\beta + \varepsilon_{ist} \quad (4)$$

In this framework, each individual has two potential outcomes: one with treatment and one without treatment. However, our observations are restricted to the outcomes corresponding to each group (treated or not treated) in t=1. In theory, these outcomes should diverge due to the presence or absence of the treatment and can be expressed as:

$$Y_{ist}(OM) = OM_{is} \times Y_{ist}(1) + (1 - OM_{is}) \times Y_{ist}(0)$$

Under the assumption that the treated group would follow a predetermined trajectory in the absence of treatment, any deviation from this path can be attributed to the causal impact of the treatment on this group. This deviation, denoted as the Average Treatment Effect on the Treated (ATT), is described in Equation 5.

$$\mathrm{ATT} = \underbrace{E(Y_{is1}(1) \backslash OM_{is} = 1)}_{A=Observerd\ outcome\ for\ treated} - \underbrace{E(Y_{is1}(0) \backslash OM_{is} = 1)}_{B=Unobserverd\ outcome\ for\ treated} \quad (5)$$

In Equation 5, we have information about the value of part A, as it represents the observed outcome for the treated group in t=1 after the treatment. However, when it comes to part B (as defined in

Equation 5), the path that the treated group would have followed in the absence of treatment is unknown. To make this estimation, we rely on the assumption that this path would be parallel to the trajectory followed by the control group. This assumption is referred to as the Parallel Trend Assumption (PTA). In simpler terms, we assume that the unobserved path taken by the treated group (B) in the scenario where they did not receive treatment is the same as the observed path in the control group (Equation 6).

$$E(Y_{is1}(0) - Y_{is0} \backslash OM_{is} = 1) = E(Y_{is1} - Y_{is0} \backslash OM_{is} = 0) \quad (6)$$

Finally, using Equation 6 in Equation 5, we can construct a feasible estimator for the ATT that will be given by:

$$\widehat{ATT} = \left[E\left(Y_{is,1} - Y_{is,0} | OM_{is} = 1\right)\right] - \left[E\left(Y_{is,1} - Y_{is,0} | OM_{is} = 0\right)\right]$$
$$= E\left(Y_{is,1} | OM_{is} = 1\right) - \hat{E}\left(Y_{is,1}(0) | OM_{i} = 1\right) \quad (7)$$

While the Parallel Trend Assumption (PTA) can be challenging to satisfy in practice due to potential dissimilarities between the treated and control groups, Callaway and Sant'Anna (2021) propose a generalized approach incorporating additional groups and fixed effects in the specification. DiD designs often involve more than two periods or treated groups, further complicating the PTA assumption. To mitigate this, Sant'Anna and Zhao (2020) suggest using the PTA for groups with identical pre-treatment characteristics (α and X), minimizing bias due to group differences. Let "W" represent the set of individuals (α) and mine-linked characteristics (X). Here, $\theta(W_{is})$ is the ΔY_{is} if there was no treatment based on pre-treatment characteristics. With this updated assumption, the new DiD estimator is $\widehat{ATT}_*$ (Equation 9).

$$E(Y_{is1}(0) - Y_{is0} \backslash OM_{is} = 1, W_{is}) = E(Y_{is1} - Y_{is0} \backslash OM_{is} = 0, W_{is}) = \theta(W) \quad (8)$$

$$\widehat{ATT}_* = E\left(Y_{is,1} | OM_{is} = 1\right) - E\left(Y_{is,0} | OM_{is} = 1\right) + \hat{E}(\theta(W_{is}) | OM_{is} = 1) \quad (9)$$

Various approaches have been proposed in the literature to estimate the component $\hat{E}(\theta(W_{is})|OM_{is}=1)$ from Equation 9. In this paper, as it is a cross section database with many different individuals, we adopt the Outcome Regression Approach (OR) estimator proposed by Sant'Anna & Zhao (2020). The OR estimator employs a two-step procedure. In the first step using data from the control group, we model $E(\theta_{is}|W_{is})=\theta(W_{is})$ as a function of W, so $E(\theta_{is}|W_{is}=w_{is})=\theta(w_{is})\ \forall\ i|OM_{is}=0$. Then, $E(\theta_{is}|OM_{is}=1)$ is estimated by substituting θ_{is} with the predicted outcome $\hat{\theta}(w_{is})$. Thus the OR estimator for the ATT becomes:

$$\widehat{ATT}_{or}=E(\Delta Y_{is}|OM_{is}=1)-E\left(\hat{\theta}(w_{is})\middle|OM_{is}=1\right) \tag{10}$$

To handle the varying timing and groups affected by mine operations, we adopt the framework proposed by Callaway and Sant'Anna (2021), building upon the work of Sant'Anna and Zhao (2020). This approach introduces a designated group "g" to represent the cohort of mine operation, accommodating temporal variations marked by "t." Equation 11 illustrates how this framework allows us to track the evolution of the proposed ATT over time within a specific group. In our estimation, we implement the methodology developed by Rios-Avila et al. (2021), which is based on Callaway and Sant'Anna (2021). This approach dissects the combinations of groups and times into multiple 2X2 models and then aggregates them based on "g." By following this approach, we can identify ATTs for each treated group "G" at every time point "t" ($ATT(g,t)$).

$$\begin{aligned}ATT(g,t)\ &=E\left(Y_{is,t}(g)-Y_{is,t}(0)\middle|G_{is}=g\right)=E\left(Y_{is,t}(g)-Y_{i,t}(0)\middle|G_{is}=g\right)\\&=E\left(Y_{is,t}\middle|G_{is}=g\right)-\left(E\left(Y_{is,g-1}|G_{is}=g\right)+E\left(Y_{is,t}(0)-Y_{is,g-1}|G_{i}=g\right)\right)\end{aligned}$$

$$ATT(g,t)=E\left(Y_{is,t}-Y_{is,g-1}\middle|G_{is}=g\right)\ -\left(E\left(Y_{is,t}(0)-Y_{is,g-1}|G_{is}=g\right)\right) \tag{11}$$

Following this process, we calculate an ATT and corresponding weights for each group within each period. This enables us to consolidate the ATT over time, similar to an event analysis as detailed in the results section, and by group to analyze the impacts within each group and make comparisons.

As previously mentioned, the groups may vary in their timing. Thus, in this framework, the population, initially divided into two groups (treatment and control), is now sorted into three sets: treated, not yet treated, and control. The final step is to estimate the expected change in outcomes in the absence of treatment. For this purpose, we apply the conditional PTA assumption to the "not yet treated" group. The PTA assumption is defined as:

$$E\big(Y_{is,t}(0) - Y_{is,g-1}|G_{is} = g\big) = E\big(Y_{is,t} - Y_{is,g-1}|G_{is} = 0\big)$$

So, Equation 12 describes the final ATT to be estimated using the PTA for not yet treated.

$$ATT(g,t) = E\big(Y_{is,t} - Y_{is,g-1}\big|G_{is} = g\big) - E\big(Y_{is,t} - Y_{is,g-1}\big|G_{is} = 0\big) \quad (12)$$

This framework allows us to estimate the causal impact of each period in which mines start to operate in the census and examine how this impact changes over time. We implement both the Simple and Event aggregation methods from Rios-Avila et al. (2021). These aggregation methods are defined as:

$$ATT_{Simple} = \frac{\sum_{t\geq g} w_{g,t} ATT(g,t)}{\sum_{t\geq g} w_{g,t}} \quad (13) \qquad ATT_{event} = \frac{\sum_{t+e=g} w_{g,t} ATT(g,t)}{\sum_{t+e=g} w_{g,t}} \quad (14)$$

Results

In this section, we first analyze the outcomes derived from the Panel Ordinary Least Squares (OLS) model (Equation 1). Then, we proceed to present the findings from the Instrumental Variables (IV) approach, specifically focusing on the Local Average Treatment Effects (LATE) estimation as outlined in Equation 3. To conclude, we analyze the results of the heterogeneous Difference-in-Differences results (ATT estimation) in accordance with Equation 13.

OLS Results

The duration of the nearest mine's operation has a positive and significant impact on secondary school cohort size and higher education enrollment. Conversely, it has a negative and significant effect on Saber 11 test scores. Interestingly, it does not significantly affect the probability of labor market enrollment (Table 2).

Moreover, schools located farther from the nearest mine tend to have smaller cohorts, lower Saber 11 test scores, and a lower probability of college enrollment. However, they do exhibit a slightly higher probability of enrolling in the labor market. Longer mine operation positively influences all four output variables of interest. Notably, having more mines within a 1km radius negatively affects cohort size, Saber 11 test scores, and college enrollment but positively impacts the probability of labor market enrollment (Table 2A).

Examining the materials extracted, CO-AS extraction has a negative influence on cohort size, Saber 11 test scores, and labor market enrollment. Conversely, Gold extraction has a positive impact on cohort size, test scores, and enrollment in higher education (Table 3).

Instrumental Variables Results

The initial phase in obtaining the NL2SLS estimator involves using instruments and regular controls from Equation 1 to estimate the probability of the nearest mine's operational status in Equation 2. Various instruments, including the price index PMETA, as well as the prices of Aluminum, Gold, Iron ore, and Zinc, were tested, and all yielded robust instruments. In Table 3A, Columns 1 to 5 present the results for Equation 2 using these instruments. PMETA was selected as it comprises a sample of metals that offers greater accuracy given the diversity of mines in the country. In the subsequent step, employing PMETA as an instrument for OM, the Equation 3 was estimated. The coefficients obtained represent the Local Average Treatment Effect (LATE) and reveal the causal impact of the mine's operation on various outcomes (Table 3).

The primary findings indicate that the coefficient for the Saber 11 test score is not statistically significant, while the probability of enrollment in the labor market is negative and significant. In contrast, the size of the cohort (6%) and the probability of enrolling in college (4.5%) are positive and significant, and their effects are stronger than those found in the OLS section.

Specifically, Co-As and Metals are associated with a 16.3% and 15.7% reduction in cohort size, respectively, while Gold, Construction, and Other mines positively impact cohort sizes by 14.4%, 5%, and 26.3%, respectively. Regarding Saber 11 test scores, there is an average increase of 3.8 points when the nearest mine extracts gold. In contrast, there are reductions of 3.6 points and 2.13 points when the nearest mine extracts metals or other materials, respectively.

Despite the decline in test scores, the probability of enrolling in college increases for students whose closest mine extracts construction products. Conversely, students closer to a mine that extracts metals experience a decrease in the probability of enrolling in college. There is no significant effect for Co-As, Gold, and Other in terms of the probability of college enrollment.

Finally, the results show that the probability of securing formal employment decreases by 2 percentage points if the closest mine to the school is operational. The negative impact is generally small, but there is significant heterogeneity across most mine types. If the closer mine extracts Co-As or Other products, the probability of enrolling in the labor market decreases by 1% and 1.6%, respectively. However, if the secondary school is located near a mine that extracts construction products, Gold, or Metals, the probability of enrolling in the labor market increases by 1.1%, 1%, and 4.6%, respectively.

Heterogeneous Difference in Differences (DiD) Results Analysis

In this subsection, we aim to analyze the Average Treatment Effect on the Treated (ATT) for the closest mine in operation, employing the heterogeneous difference-in-differences (DiD) results from Equation 13 using the CSDID command by Rios-Avila et al. (2021).

To support our results, we rely on the Parallel Trend Assumption (PTA), which requires the Pre-treatment average to be nonsignificant. In cases where the pre-treatment differs from zero, we need a significant change in trend (inverting the sign with significant values) after the treatment, along with other tests to support the results. In this case, the results for the Saber 11 test score and the enrollment in the labor market hold the PTA test, allowing us to rely on the reported ATT.

Regarding the Size of the cohort, the ATT reports an increase of about 7.2%, slightly higher than the 6% reported by the IV approach and the 3.5% from the OLS. The Saber 11 test score shows an increase of 3.8 points (impact of 8.2% compared with the control mean), which is positive and significant, differing from the result from LATE and with the opposite sign from the report from the OLS (Table 2).

The probability of enrolling in college increases by 12.2% according to the ATT if the closer mine starts to operate. This coefficient is positive and significant, aligning with the positive results from the IV approach and the OLS. The ATT approach reports a reduction of 8.6% in the probability of enrolling in the labor market. This result is negative, similar to the LATE approach result, but also stronger than the LATE coefficient. we also examined whether the impact of the closer mine was specific to the mining sector, and we found similar results in the ATT and LATE, reporting a decrease in the probability of enrolling in the labor market in the mining sector of 1 percentage point (Table 2).

Finally, as the Saber 11 test score and Enrollment in the Labor market satisfy the PTA assumption, the aggregation using Equation 14 enables us to investigate the impact of the closest mine in operation, resembling an event study. Figure 3 illustrates that although the increase in the Saber 11 test score averages about 3.8 points, it actually peaks at 17.9 points in the ninth year after the mine starts operating. The average for the post-treatment period is 7.28 points. Figure 4 presents the time event for the probability of enrollment in the labor market. In this case, there is a decline of 25.7%

in the ninth year after the mine commences operation, and the full post-treatment period records a decrease in the probability of labor market enrollment of 13%.

Conclusions

The key findings of this study highlight significant disparities across different types of mined products and their impacts on various outcomes. Consistent with Angrist & Kugler (2008), the observed increase in student cohort size of about 6% is notable. The positive influence of the nearest operational mine on cohort size is evident across all three analyzed approaches. However, schools in proximity to Gold or Co-As and Metals extraction mines show a notable decrease in cohort size. Possible reasons for this decline include the establishment of new schools, student migration due to contamination concerns, or the emergence of informal businesses drawing students away from academics. Wood (2009)'s framework suggests that students with lower academic performance may discontinue their education due to high opportunity costs, especially if the mine encourages informal employment.

While the study revealed no major changes in academic performance, as evidenced by higher Saber 11 test scores, it showed that legal Gold mines tend to improve the Saber 11 test score, contrary to the findings of Bonilla Mejía (2020) for both legal and illegal Gold mines, whereas other metal extractions have negative impacts. It is important to highlight that the effect of an operational mine nearby is positive, resulting in an increased cohort size annually without compromising academic performance (as measured by the Saber 11 test score) or even enhancing the quality of education (as measured by the increased probability of college enrollment). This effect occurs even as the demand for education outpaces the fixed supply, which is particularly relevant as establishing a new school requires time. Remarkably, the system has effectively managed potential issues related to overcrowding in cohorts, leading to a rise in the participation of secondary graduates in college, and subsequently reducing the likelihood of immediate entry into the formal labor market. These consistent trends suggest that students are actively choosing to prioritize their continued education over immediate employment, underscoring the positive impact of the mine on educational aspirations.

Additionally, the distance from the mine significantly affects cohort size, Saber 11 test scores, and the probability of college enrollment, with a positive effect on labor market participation. Although the size of the mine has a significant but minimal influence, the type of extracted product plays a crucial role. Notably, Gold mining has been shown to increase student cohort sizes and Saber 11 scores without affecting the likelihood of college attendance, a result similar to what is found in schools closer to mines extracting Other products. These findings contrast with the effects observed in schools closer to mines extracting Construction materials, where an increase in cohort size affects the Saber 11 score but not the probability of college enrollment or labor market participation.

Ultimately, this research serves as a vital tool for policymakers grappling with the intricate balance between the economic benefits of mining operations and the imperative of sustainable resource management. It underscores the need for effective regulations and enforcement measures in the extractive industries, safeguarding both the environment and the long-term well-being of Colombian citizens. Proper regulation and enforcement mechanisms are vital to ensure that mining operations remain sustainable, avoid illegal practices, and contribute positively to local communities, preserving the life path of the nation's young students.

Acknowledgements: All views and errors are ours. The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

This paper utilizes confidential administrative data from ICFES and SPADIES (Ministry of Education of Colombia), and PILA (Ministry of Health of Colombia). Access to the data can be obtained by a request letter to each Ministry, and the authors are available to assist in facilitating this contact.

References:

Agüero, J. M., Balcázar, C. F., Maldonado, S., & Ñopo, H. (2016). Natural Resources, Redistribution and Human Capital Formation. *Documentos de Trabajo Universidad Del Rosario*, *1–46*.

Almond, D., Edlund, L., & Marten, P. (2009). Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *Quarterly Journal of Economics*, *124*(4), 1729–1772.

Álvarez, R., & Vergara, D. (2022). Natural resources and educational attainment: Evidence from Chile. *Resources Policy*, *76*, 102573. https://doi.org/10.1016/J.RESOURPOL.2022.102573

Angrist, J. D., & Kugler, A. (2008). Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia. *Review of Economics and Statistics*, 1–53. https://doi.org/10.1162/rest.90.2.191

Balza, L. H., De los Rios, C., Jimenez, R., & Manzano, O. (2021). *The Local Human Capital Cost of Oil Exploitation* (1258; IDB-WP, Issue July).

Barro, R. J. (2001). Human capital: Growth, history, and policy - A session to honor Stanley Engerman: Human capital and growth. *American Economic Review*, *91*(2), 12–17.

Black, D. A., McKinnish, T. G., & Sanders, S. (2005). *No Title*. *115*(503), 449–476. https://doi.org/10.1111/J.1468-0297.2005.00996.X

Bonilla Mejía, L. (2020). Mining and human capital accumulation: Evidence from the Colombian gold rush. *Journal of Development Economics*, *145*, 102471. https://doi.org/10.1016/j.jdeveco.2020.102471

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Caselli, F. (2006). Power Struggles and the Natural Resource Curse. *LSE Working Paper*, *December 2005*, 1–2.

Caselli, F., & Michaels, G. (2013). Do oil windfalls improve living standards? Evidence from brazil. *American Economic Journal: Applied Economics*, *5*(1), 208–238. https://doi.org/10.1257/app.5.1.208

Currie, J., Hanushek, E., Kahn, E. M., Neidell, M., & Rivkin, S. G. (2009). Does Pollution Increase School Absences? *Review of Economics and Statistics*, *91*(4), 682–694. https://doi.org/10.1162/rest.91.4.682

Currie, J., Zivin, J. G., Mullins, J., & Neidell, M. (2014). What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution? *Annual Review of Resource Economics*, *6*(1), 217–247. https://doi.org/10.1146/annurev-resource-100913-012610

Dube, O., & Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies*, *80*(4), 1384–1421. https://doi.org/10.1093/restud/rdt009

Gilliland, F., Berhane, K., Rappaport, E., Thomas, D., Avol, E., Gauderman, J., London, S., Margolis, H., McConnell, R., Islam, K. T., & Peters, J. (2001). The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology (Cambridge, Mass.)*, *12*(1), 43–54. https://doi.org/10.1097/00001648-200101000-00009

Gylfason, T. (2001). Natural resources, education, and economic development. *European Economic Review*, *45*(4–6), 847–859. https://doi.org/10.1016/S0014-2921(01)00127-1

Hanushek, E., & Woessmann, L. (2010). Education and Economic Growth Early Studies of Schooling Quantity and Economic Growth. *International Encyclopedia of Education*, *2*, 245–252.

International Monetary Fund. (2023). *IMF Primary Commodity Prices*. https://www.imf.org/en/Research/commodity-prices

Lavy, V., & Roth, S. (2014). The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation. *National Bureau of Economic Research*, 1–40. https://doi.org/10.3386/w20648

Litschig, S., & Morrison, K. M. (2013). The impact of intergovernmental transfers on education outcomes and poverty reduction. *American Economic Journal: Applied Economics*, *5*(4), 206–240. https://doi.org/10.1257/app.5.4.206

Loayza, N., & Rigolini, J. (2016). The Local Impact of Mining on Poverty and Inequality : Evidence from the Commodity Boom in Peru The Local Impact of Mining on Poverty and Inequality : Evidence from the Commodity Boom in Peru. *World Development*, *84*(33), 219–234. https://www.sciencedirect.com/science/article/abs/pii/S0305750X16000462

Loayza, N., Rigolini, J., & Calvo-González, O. (2014). More than you can handle: Decentralization and spending ability of peruvian municipalities. *Economics and Politics*, *26*(1), 56–78. https://doi.org/10.1111/ecpo.12026

Maldonado, S. (2018). *The Non-Monotonic Political Effects of Resource Booms* (121).

Maldonado, S., & Ardanaz, M. (2023). Natural resource windfalls and efficiency in local government expenditure: Evidence from Peru. *Economics and Politics*, *35*(1), 28–64. https://doi.org/10.1111/ECPO.12209

Martínez, L. R. (2023). Natural Resource Rents, Local Taxes, and Government Performance: Evidence from Colombia. *Review of Economics and Statistics*, 1–28. https://doi.org/10.1162/rest_a_01324

Martínez Ortiz, A., & Aguilar, T. (2012). Impacto socioeconómico de la Minería en Colombia. In *Sector de Minería a Gran Escala*. https://www.repository.fedesarrollo.org.co/bitstream/handle/11445/375/Impacto socioeconomico de la mineria en Colombia - Informe_Impacto_de_la_mineria_Final 26 abril.pdf?sequence=2&isAllowed=y#:~:text=Dentro de los aspectos positivos,peso en las finanzas

Martínez, Z., González, M., Paternina, J., & Cantero, M. (2017). Crop soils pollution by heavy metals, the Alacran mining area, Córdoba-Colombia. *Revista Temas Agrarios*, *22*(2), 20–32.

Michaels, G. (2011). The Long Term Consequences of Resource-Based Specialisation*. *The Economic Journal*, *121*(551), 31–57. https://doi.org/10.1111/J.1468-0297.2010.02402.X

Park, H., Lee, B., Ha, E.-H., Lee, J.-T., Kim, H., & Hong, Y.-C. (2002). Association of Air Pollution With School Absenteeism Due to Illness. *Archives of Pediatrics & Adolescent Medicine*, *156*(12), 1235–1239. https://doi.org/10.1001/archpedi.156.12.1235

Pegg, S. (2006). Mining and poverty reduction: Transforming rhetoric into reality. *Journal of Cleaner Production*, *14*(3–4), 376–387. https://doi.org/10.1016/j.jclepro.2004.06.006

Romero, M., & Saavedra, S. (2016). *The Effects of Gold Mining on Newborns' Health*.

Ross, M. (2015). What Have We Learned about the Resource Curse? *Annual Review of Political Science*, *18*(1), 239–259. https://doi.org/10.1146/annurev-polisci-052213-040359

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/J.JECONOM.2020.06.003

Torvik, R. (2002). Natural resources, rent seeking and welfare. *Journal of Development Economics*, *67*(2), 455–470. https://doi.org/10.1016/S0304-3878(01)00195-X

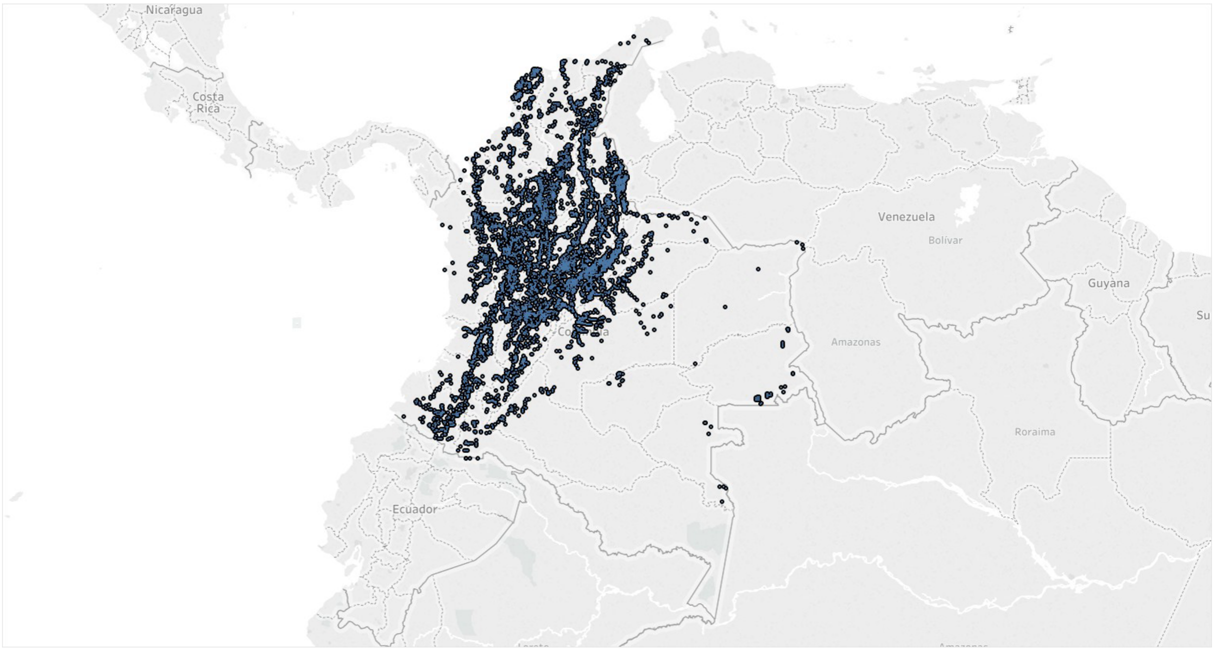
van der Ploeg, F. (2011). Natural Resources: Curse or Blessing? *Journal of Economic Literature*, *49*(2), 366–420. https://doi.org/10.1257/jel.49.2.366

Wood, T. (2009). The Sheepskin Effect. In *National Association of Scholars* (Issue July). https://www.nas.org/blogs/article/the_sheepskin_effect

.

Figures

Figure 1. Location of legal mines in Colombia



Note: The map shows the location of the mining titles according to the geolocation made by Activisual. We downloaded the Colombian mining census data for 2014 from Tierra Minera in 2017.

Figure 2. Location of secondary schools in Colombia

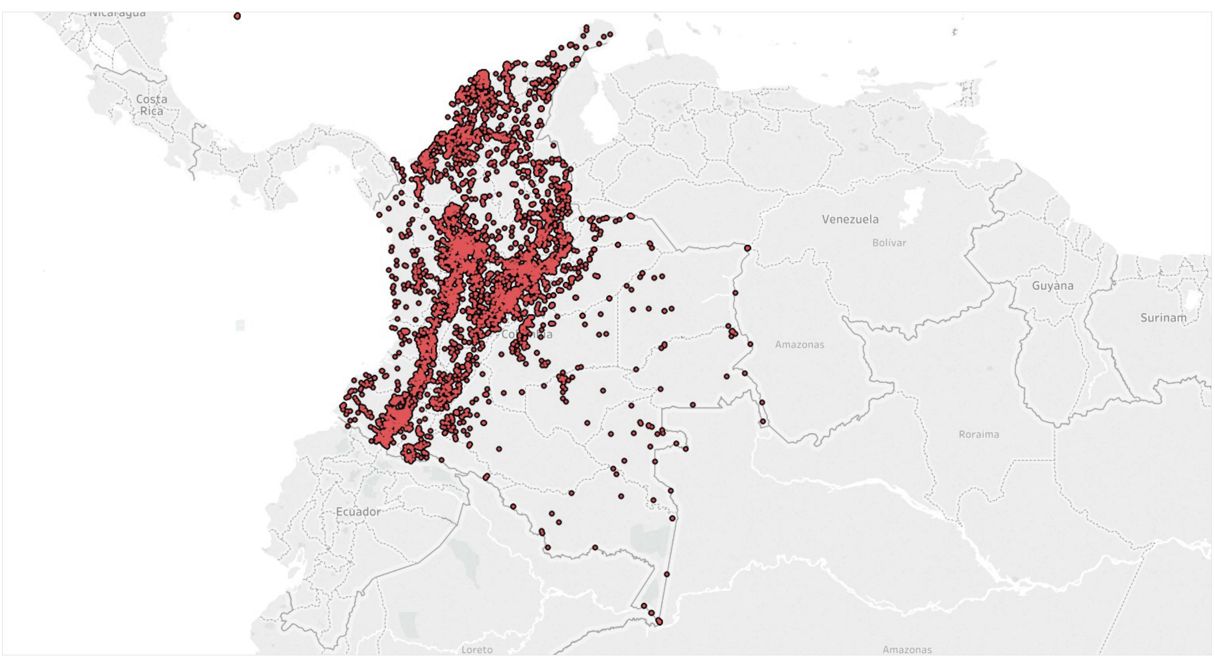
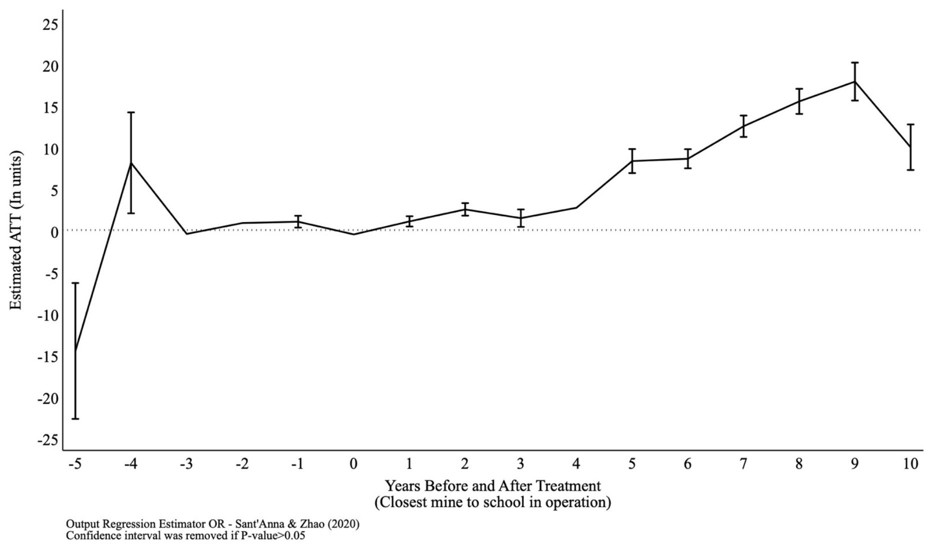
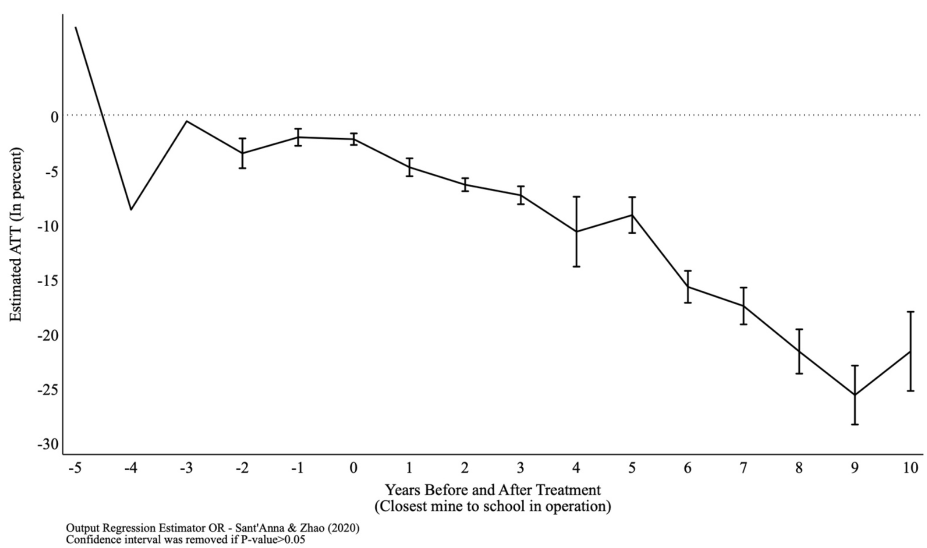


Figure 3. ATT – Saber 11 test score



Notes: The figure shows the estimated ATT coefficients for the Saber 11 test score. The treated group consists of individuals who presented the Saber 11 test while they were enrolled in a secondary school whose closest mine was not in operation at the moment of the exam (dotted horizontal line in Y=0). The X-axis represents the years before and after the closest mine starts operation, while the whiskers depict the 95 percent confidence intervals. Confidence interval was removed if Pvalue>0.05. The coefficients are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 14) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) OR estimator (Equation 10). For further details on the variables used as controls, readers can refer to Table 1.

Figure 4. ATT – Enrollment in Labor Market



Notes: The figure shows the estimated ATT coefficients for the probability of enrollment in the labor market. The treated group consists of individuals who presented the Saber 11 test while they were enrolled in a secondary school whose closest mine was not in operation at the moment of the exam (dotted horizontal line in Y=0). The X-axis represents the years before and after the closest mine starts operation, while the whiskers depict the 95 percent confidence intervals. Confidence interval was removed if Pvalue>0.05. The coefficients are obtained utilizing the framework proposed by Callaway and Sant'Anna (2021) (Equation 14) through the Rios-Avila et al. (2021) methodology for the Sant'Anna and Zhao (2020) OR estimator (Equation 10). For further details on the variables used as controls, readers can refer to Table 1.

Tables

Table 1. Descriptive Statistics

Variables	(1) mean	(2) sd	(3) min	(4) max
Size of cohort (in units)	101.5	104.8	1	1,316
Size of cohort (in log)	4.3	0.8	0	7
Saber 11 score	50.7	28.9	1	100
Enrolled in higher education	50.6	50.0	0	100
Enrolled in formal labor market	8.0	27.2	0	100
Enrolled in formal labor market (mining sector)	0.1	2.8	0	100
PMETA (IMF metals index)	135.1	51.7	41.3	209
Closest mine is operating	67.9	46.7	0	100
Distance to closest mine (in km)	7.1	47	0.03	2,899
$Distance^2$	2,262	97,644	0.001	8,406,000
Mine operation time (in years)	5.0	8.1	-12	23
Number of mines in a ratio of 1 km	0.1	0.7	0	35
Number of mines in a ratio of 3 km	2.0	4.5	0	115
Number of mines in a ratio of 5 km	5.9	10.2	0	128
Number of mines in a ratio of 10 km	22.6	27.5	0	250
Number of mines in a ratio of 25 km	95.9	81.1	0	565
Number of mines in a ratio of 50 km	250.9	170.0	0	900
Year of birth	1,990	5.1	1,950	2,000
Female	54.2	49.8	0	100
Public school	71.0	45.4	0	100
Household income	2.0	1.1	0	9
Father's years of education	9.5	3.8	0	17
Mother's years of education	9.6	3.7	0	17
Ethnicity group	5.5	22.8	0	100
Coed high school	94.9	22.1	0	100
Urban high school	76.5	42.4	0	100
Academic degree	52.7	49.9	0	100
School calendar from January to December	96.8	17.5	0	100
Size of the mine (in Ha)	374	4172	0	205,888
School latitude	5.97	2.62	-4.22	23.75
School longitude	-74.84	1.32	-99.11	-65.87
Observations	5,710,986			

Note: Table shows the mean, standard deviation, minimum, and maximum for the main characteristics of all secondary school graduates who took Saber 11 test from 2002 to 2012. Dummies in percent.

Table 2. Main Results

	Variables	(1) Size of cohort (in log)	(2) Saber 11 score	(3) Enrolled in higher education	(4) Enrolled in formal labor market	(5) Enrolled in formal labor market (mining sector)
Panel A. OLS	Closest mine is operating	0.035***	-0.113***	0.011***	0.000	-0.000***
		(0.001)	(0.033)	(0.001)	(0.000)	(0.000)
Panel B. IV approach	$\widehat{\text{Closest mine is operating}}$ $(LATE)$	0.060***	0.011	0.045***	-0.002***	-0.001***
		(0.002)	(0.055)	(0.001)	(0.001)	(0.000)
Panel C. DID approach	Closest mine is operating (ATT)	0.072***	3.871***	0.122***	-0.086***	-0.001*
		(0.010)	(0.371)	(0.007)	(0.001)	(0.000)
	Pre-Treatment (avg)	0.349***	0.644	0.055***	-0.006	0.002
		(0.023)	(0.481)	(0.009)	(0.007)	(0.001)
	Control mean	85.25 (in units)	47.17	0.4914	0.1	0.001
	Observations			5,710,986		

Note: Table shows the coefficients of interest for Size of cohort, Saber 11 test score, probability of enrollment in Higher Education, probability of enrollment in the labor market and in the mining sector of the labor market. Panel A estimated following specification in Equation 1, Panel B following specification in Equation 3 (first step of IV approach can be found in the Appendix). Panel C estimated with specification in Equation 12 for ATT and in Equation 14 for pre period for pretreatment check. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 3. Results with Extracted Product

	Full sample	By extracted product				
		Co-As	Construction	Gold	Metals	Other
Variables	(1)	(2)	(3)	(4)	(5)	(6)
		Panel A. OLS				
Size of cohort (in log)	0.035***	-0.065***	0.040***	0.064***	0.003	0.185***
	(0.001)	(0.005)	(0.001)	(0.004)	(0.014)	(0.004)
Saber 11 score	-0.174***	-0.974***	-0.088**	0.608***	-0.315	-0.060
	(0.033)	(0.205)	(0.037)	(0.140)	(0.456)	(0.151)
Enrolled in higher education	0.011***	0.004	0.011***	0.006**	-0.080***	0.007**
	(0.001)	(0.004)	(0.001)	(0.003)	(0.009)	(0.003)
Enrolled in formal labor market	0.000	-0.006***	0.002***	0.002	0.016***	-0.015***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.005)	(0.001)
		Panel B IV approach				
Size of cohort (in log)	0.060***	-0.163***	0.050***	0.144***	-0.157***	0.263***
	(0.002)	(0.007)	(0.002)	(0.007)	(0.023)	(0.007)
Saber 11 score	0.011	0.671**	-0.309***	3.833***	-3.594***	-2.133***
	(0.055)	(0.306)	(0.061)	(0.268)	(0.721)	(0.252)
Enrolled in higher education	0.045***	0.004	0.036***	0.007	-0.115***	0.007
	(0.001)	(0.006)	(0.001)	(0.005)	(0.014)	(0.005)
Enrolled in formal labor market	-0.002***	-0.010***	0.011***	0.010***	0.046***	-0.016***
	(0.001)	(0.003)	(0.001)	(0.003)	(0.008)	(0.002)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906

Note: Table shows the coefficients of interest for Size of cohort, Saber 11 test score, probability of enrollment in Higher Education, probability of enrollment in the labor market and in the mining sector of the labor market. Panel A estimated following specification in Equation 1 (Full results can be found in the Appendix), Panel B following specification in Equation 3 (first step of IV approach and full regression results can be found in the Appendix). Co-As is Coal and Asbestos- Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Appendix

Table 1A. Status of mines according to census 2014

Department	In Progress	Reactivated	Discontinued	Total
Antioquia	1,580	8	1	1,589
Arauca	44	0	0	44
Atlántico	98	2	0	100
Bogota	34	0	0	34
Bolivar	444	0	0	444
Boyacá	1,534	3	1	1,538
Caldas	396	1	1	398
Caquetá	60	0	0	60
Casanare	147	2	0	149
Cauca	216	0	0	216
Cesar	382	0	2	384
Chocó	162	0	16	178
Córdoba	103	0	1	104
Cundinamarca	993	4	0	997
Guainía	32	2	0	34
La Guajira	53	0	0	53
Guaviare	14	0	0	14
Huila	211	0	0	211
Magdalena	73	1	0	74
Meta	226	0	0	226
Nariño	206	0	0	206
N.Santander	709	0	0	709
Putumayo	52	0	0	52
Quindío	68	0	0	68
Risaralda	69	0	0	69
Santander	673	1	0	674
Sucre	67	0	0	67
Tolima	595	0	2	597
Valle Del Cauca	304	0	0	304
Vaupes	9	0	0	9
Vichada	6	0	0	6
Total	9,545	24	24	9,593

Source: Agencia Nacional de Mineria (National Agency for Mining) -ANM-; Tierra Minera

Table 2A. First Step. Marginal Effects

	A. Selected Index	B. Other ores price indexes				
	PMETA	Aluminum	Copper	Gold	Iron	Zinc
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Closest mine is operating	Closest mine is operating	Closest mine is operating	Closest mine is operating	Closest mine is operating	Closest mine is operating
Price Index x (Mines 1 km +1)	-0.000***	0.000***	-0.000***	-0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to closest mine (in km)	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Distance^2$	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.032***	0.032***	0.032***	0.032***	0.032***	0.032***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 1 km	-0.000	-0.012***	-0.001***	0.001*	-0.002***	-0.011***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Number of mines in a ratio of 3 km	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 5 km	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 10 km	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 25 km	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Public school	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household income	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Father's years of education	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mother's years of education	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Coed high school	-0.032***	-0.031***	-0.032***	-0.032***	-0.032***	-0.031***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Urban high school	0.026***	0.026***	0.026***	0.026***	0.026***	0.026***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Academic degree	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School calendar from Jan to Dec	-0.012***	-0.012***	-0.012***	-0.012***	-0.012***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Size of the mine	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	0.019***	0.019***	0.019***	0.019***	0.019***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School longitude	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	5,710,986	5,710,986	5,710,986	5,710,986	5,710,986	5,710,986
$PS\ R^2$	0.520	0.520	0.520	0.520	0.520	0.520
$Chi^2\ p-value$	0	0	0	0	0	0

Note: The table displays the coefficients of interest for the first step in the IV approach (Equation 2). Panel A presents the results for PMETA, while Panel B divides the results by other types of commodities. The estimations follow the specifications outlined in Equation 2 with a Probit model, with the regression incorporating time and departmental controls (not shown). Marginal effects for the output are shown. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 3A. Size of cohort (OLS)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)
Closest mine is operating	0.035***	-0.065***	0.040***	0.064***	0.003	0.185***
	(0.001)	(0.005)	(0.001)	(0.004)	(0.014)	(0.004)
Distance to closest mine (in km)	-0.000***	0.019***	-0.001***	0.000*	0.028***	-0.011***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000**	-0.001***	-0.000***	0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.007***	0.018***	0.005***	-0.003***	-0.036***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)
Number of mines in a ratio of 1 km	-0.046***	-0.009***	-0.057***	0.000	0.378***	0.046***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.029)	(0.004)
Number of mines in a ratio of 3 km	-0.005***	0.003***	-0.002***	-0.031***	-0.388***	-0.146***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.012)	(0.002)
Number of mines in a ratio of 5 km	0.002***	0.001	0.001***	0.031***	0.065***	0.054***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)
Number of mines in a ratio of 10 km	0.001***	0.000**	0.001***	-0.003***	0.024***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 25 km	0.000***	-0.002***	0.001***	-0.001***	-0.007***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	0.000***	0.001***	0.000***	-0.000***	0.004***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	0.001***	0.004***	0.000*	0.003***	-0.018***	0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Female	-0.005***	0.006*	-0.007***	0.029***	0.018**	-0.020***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.008)	(0.003)
Public school	0.418***	0.733***	0.418***	0.371***	0.325***	0.350***
	(0.001)	(0.005)	(0.001)	(0.005)	(0.024)	(0.004)
Household income	0.040***	0.057***	0.033***	0.081***	-0.024***	0.047***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.005)	(0.001)
Father's years of education	0.008***	0.006***	0.008***	0.009***	-0.006***	0.013***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Mother's years of education	0.008***	0.006***	0.007***	0.015***	-0.015***	0.014***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Ethnicity	0.089***	0.069***	0.073***	-0.148***		0.287***
	(0.002)	(0.006)	(0.002)	(0.011)		(0.006)
Coed high school	-0.241***	0.024**	-0.267***	-0.189***		0.202***
	(0.001)	(0.011)	(0.002)	(0.010)		(0.006)
Urban high school	0.507***	0.477***	0.476***	0.540***	0.511***	0.411***
	(0.001)	(0.005)	(0.001)	(0.003)	(0.023)	(0.004)
Academic degree	-0.194***	-0.180***	-0.196***	-0.148***	-0.169***	-0.242***
	(0.001)	(0.004)	(0.001)	(0.003)	(0.010)	(0.004)
School calendar from Jan to Dec	0.145***	-0.144***	0.125***	-0.085***		0.416***
	(0.002)	(0.008)	(0.002)	(0.009)		(0.011)
Size of the mine	-0.000***	-0.000***	0.000***	-0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	0.040***	0.025***	0.031***	-0.057***	1.079***	0.095***
	(0.001)	(0.005)	(0.001)	(0.003)	(0.032)	(0.005)
School longitude	-0.014***	0.187***	-0.016***	0.067***	0.247***	0.044***
	(0.001)	(0.006)	(0.001)	(0.004)	(0.024)	(0.007)
Constant	-0.031	7.406***	1.275***	2.345***	52.861***	-14.938***
	(0.180)	(0.920)	(0.199)	(0.685)	(2.719)	(0.895)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.178	0.264	0.170	0.242	0.340	0.313

Note: The table displays the coefficients of interest for the Size of the cohort. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 1, with the regression incorporating time and departmental controls (not shown). Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 4A. Saber 11 test score (OLS)

Variables	A. Full sample (1) Saber 11 test score	B. By extracted product Co-As (2) Saber 11 test score	Construction (3) Saber 11 test score	Gold (4) Saber 11 test score	Metals (5) Saber 11 test score	Other (6) Saber 11 test score
Closest mine is operating	-0.174***	-0.974***	-0.088**	0.608***	-0.315	-0.060
	(0.033)	(0.205)	(0.037)	(0.140)	(0.456)	(0.151)
Distance to closest mine (in km)	-0.045***	0.038	-0.175***	-0.137***	-0.203***	0.136***
	(0.001)	(0.042)	(0.003)	(0.008)	(0.057)	(0.015)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000***	-0.000	0.000***	0.000***	0.000***	-0.000***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.076***	0.151***	0.046***	-0.049***	-0.301***	0.045***
	(0.002)	(0.018)	(0.002)	(0.014)	(0.110)	(0.014)
Number of mines in a ratio of 1 km	-0.410***	-0.574***	-0.301***	0.010	-4.147***	-1.063***
	(0.018)	(0.078)	(0.022)	(0.047)	(0.937)	(0.146)
Number of mines in a ratio of 3 km	0.068***	0.034	0.024***	0.183***	0.566	0.607***
	(0.005)	(0.036)	(0.006)	(0.018)	(0.378)	(0.055)
Number of mines in a ratio of 5 km	0.045***	0.041*	0.064***	-0.257***	-0.364***	-0.081***
	(0.003)	(0.022)	(0.003)	(0.015)	(0.106)	(0.027)
Number of mines in a ratio of 10 km	-0.053***	0.021***	-0.069***	-0.041***	-0.012	-0.025***
	(0.001)	(0.007)	(0.001)	(0.005)	(0.037)	(0.005)
Number of mines in a ratio of 25 km	0.003***	-0.013***	0.003***	0.022***	0.040***	0.010***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.008)	(0.002)
Number of mines in a ratio of 50 km	0.012***	0.004***	0.011***	0.017***	-0.024***	0.008***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.004)	(0.001)
Year of birth	0.637***	0.521***	0.639***	0.557***	0.489***	0.637***
	(0.003)	(0.016)	(0.003)	(0.012)	(0.030)	(0.013)
Female	-4.122***	-3.565***	-4.182***	-3.528***	-3.898***	-3.858***
	(0.022)	(0.119)	(0.024)	(0.089)	(0.251)	(0.096)
Public school	-1.851***	-0.912***	-1.584***	-2.897***	6.623***	-6.632***
	(0.030)	(0.206)	(0.032)	(0.177)	(0.786)	(0.162)
Household income	4.553***	4.098***	4.538***	3.839***	2.956***	4.046***
	(0.012)	(0.070)	(0.014)	(0.053)	(0.174)	(0.053)
Father's years of education	0.670***	0.497***	0.675***	0.472***	0.683***	0.732***
	(0.005)	(0.029)	(0.006)	(0.020)	(0.061)	(0.024)
Mother's years of education	0.871***	0.672***	0.874***	0.784***	1.197***	0.929***
	(0.006)	(0.031)	(0.006)	(0.022)	(0.068)	(0.026)
Ethnicity	-0.940***	-2.943***	-0.605***	-7.710***		-0.035
	(0.054)	(0.232)	(0.059)	(0.407)		(0.206)
Coed high school	-13.858***	-9.042***	-13.796***	-6.968***		-15.306***
	(0.052)	(0.461)	(0.055)	(0.357)		(0.221)
Urban high school	1.605***	3.714***	1.881***	-0.020	-0.389	-0.341**
	(0.031)	(0.187)	(0.035)	(0.110)	(0.731)	(0.153)
Academic degree	-0.865***	-0.675***	-0.959***	-1.417***	-2.370***	0.686***
	(0.026)	(0.150)	(0.029)	(0.104)	(0.332)	(0.128)
School calendar from Jan to Dec	-4.472***	-7.899***	-3.953***	-2.788***		-8.231***
	(0.071)	(0.311)	(0.077)	(0.346)		(0.404)
Size of the mine	-0.000***	-0.000	-0.002***	-0.000***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	-0.952***	-2.018***	-1.563***	-3.888***	-1.945*	2.044***
	(0.027)	(0.194)	(0.032)	(0.113)	(1.046)	(0.172)
School longitude	2.424***	1.945***	2.446***	0.670***	-4.123***	3.868***
	(0.035)	(0.253)	(0.043)	(0.144)	(0.777)	(0.267)
Constant	-1,043.409***	-848.206***	-1,038.699***	-994.499***	-1,203.749***	-944.734***
	(6.317)	(37.606)	(7.012)	(25.661)	(88.077)	(32.382)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.161	0.136	0.158	0.223	0.135	0.175

Note: The table displays the coefficients of interest for the Saber 11 test score. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 1, with the regression incorporating time and departmental controls (not shown). Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 5A.Enrollment in Higher Education (OLS)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education
Closest mine is operating	0.011***	0.004	0.011***	0.006**	-0.080***	0.007**
	(0.001)	(0.004)	(0.001)	(0.003)	(0.009)	(0.003)
Distance to closest mine (in km)	-0.000***	0.002**	-0.000	-0.001***	0.000	-0.001***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000***	-0.000***	-0.000***	0.000**	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.000***	-0.000	0.000	-0.001***	0.012***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
Number of mines in a ratio of 1 km	-0.004***	0.001	-0.003***	-0.002*	0.046***	-0.002
	(0.000)	(0.001)	(0.000)	(0.001)	(0.018)	(0.003)
Number of mines in a ratio of 3 km	-0.000*	-0.001	-0.000***	0.000	-0.035***	-0.005***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.007)	(0.001)
Number of mines in a ratio of 5 km	0.001***	0.000	0.001***	-0.000	0.001	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
Number of mines in a ratio of 10 km	-0.000***	0.000	-0.000***	-0.000	0.003***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 25 km	0.000	-0.000***	0.000	-0.000	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	0.000***	0.000	0.000***	0.000***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	0.026***	0.025***	0.026***	0.026***	0.025***	0.028***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Female	-0.015***	-0.004*	-0.015***	-0.008***	-0.007	-0.015***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.005)	(0.002)
Public school	-0.055***	-0.059***	-0.054***	-0.060***	0.015	-0.088***
	(0.001)	(0.004)	(0.001)	(0.003)	(0.015)	(0.003)
Household income	0.039***	0.048***	0.038***	0.036***	0.041***	0.034***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)
Father's years of education	0.003***	0.002***	0.003***	0.003***	0.002	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Mother's years of education	0.005***	0.005***	0.005***	0.005***	0.006***	0.005***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Ethnicity	-0.011***	-0.034***	-0.010***	-0.079***		0.002
	(0.001)	(0.004)	(0.001)	(0.008)		(0.004)
Coed high school	-0.099***	-0.129***	-0.099***	-0.023***		-0.095***
	(0.001)	(0.009)	(0.001)	(0.007)		(0.004)
Urban high school	0.032***	0.034***	0.035***	0.023***	-0.014	0.015***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.014)	(0.003)
Academic degree	-0.008***	0.006**	-0.009***	-0.018***	-0.028***	-0.013***
	(0.000)	(0.003)	(0.001)	(0.002)	(0.006)	(0.002)
School calendar from Jan to Dec	-0.102***	-0.112***	-0.100***	-0.090***		-0.145***
	(0.001)	(0.006)	(0.001)	(0.007)		(0.007)
Size of the mine	-0.000**	-0.000***	-0.000***	0.000***	0.000*	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	0.004***	-0.004	0.003***	-0.013***	-0.111***	0.009***
	(0.000)	(0.004)	(0.001)	(0.002)	(0.020)	(0.003)
School longitude	0.004***	0.010**	0.006***	-0.020***	0.004	0.025***
	(0.001)	(0.005)	(0.001)	(0.003)	(0.015)	(0.005)
Constant	-51.166***	-48.859***	-50.746***	-51.636***	-49.968***	-52.724***
	(0.114)	(0.696)	(0.127)	(0.488)	(1.646)	(0.587)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.079	0.081	0.079	0.084	0.093	0.085

Note: The table displays the coefficients of interest for the probability of enrollment in higher education. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 1, with the regression incorporating time and departmental controls (not shown). Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 6A. Enrollment in labor market (OLS)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market
Closest mine is operating	0.000	-0.006***	0.002***	0.002	0.016***	-0.015***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.005)	(0.001)
Distance to closest mine (in km)	0.000**	-0.001***	0.000***	0.000	-0.001**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000	0.000**	-0.000***	-0.000	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.000***	0.000**	0.000***	0.000	-0.006***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 1 km	0.002***	-0.001	0.002***	0.001	0.003	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.010)	(0.001)
Number of mines in a ratio of 3 km	-0.001***	-0.001***	-0.001***	-0.000	0.007*	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.001)
Number of mines in a ratio of 5 km	-0.000	0.001***	-0.000	-0.000	0.004***	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 10 km	0.000**	-0.000***	0.000***	0.000	-0.001***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 25 km	0.000***	0.000	0.000***	0.000	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	-0.000**	-0.000*	-0.000**	-0.000***	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	-0.020***	-0.019***	-0.021***	-0.018***	-0.017***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.023***	-0.030***	-0.022***	-0.027***	-0.037***	-0.022***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)
Public school	-0.001**	0.006***	-0.001***	0.003	0.014*	0.002
	(0.000)	(0.002)	(0.000)	(0.002)	(0.008)	(0.002)
Household income	-0.003***	-0.003***	-0.002***	-0.004***	-0.005***	-0.001**
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.000)
Father's years of education	0.000**	0.000	0.000*	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Mother's years of education	-0.000***	0.000	-0.000***	-0.000	0.001	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Ethnicity	-0.001	0.001	-0.000	-0.003		-0.004**
	(0.001)	(0.002)	(0.001)	(0.004)		(0.002)
Coed high school	0.020***	0.011**	0.021***	-0.001		0.013***
	(0.001)	(0.004)	(0.001)	(0.004)		(0.002)
Urban high school	-0.001***	-0.001	-0.001**	0.002**	0.020***	-0.006***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.008)	(0.001)
Academic degree	0.003***	0.007***	0.004***	0.003***	0.004	-0.002
	(0.000)	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)
School calendar from Jan to Dec	-0.010***	0.003	-0.010***	-0.027***		-0.012***
	(0.001)	(0.003)	(0.001)	(0.004)		(0.004)
Size of the mine	-0.000	0.000	-0.000*	-0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	-0.003***	0.001	-0.003***	-0.002**	0.018*	0.001
	(0.000)	(0.002)	(0.000)	(0.001)	(0.011)	(0.002)
School longitude	0.001***	0.008***	0.003***	-0.007***	0.002	0.005*
	(0.000)	(0.002)	(0.000)	(0.001)	(0.008)	(0.002)
Constant	40.894***	38.361***	41.581***	35.308***	34.855***	39.192***
	(0.061)	(0.364)	(0.068)	(0.265)	(0.923)	(0.300)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.117	0.104	0.120	0.105	0.115	0.104

Note: The table displays the coefficients of interest for the probability of enrollment in the labor market. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 1, with the regression incorporating time and departmental controls (not shown). Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 7A. Size of cohort (IV)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)	Size of cohort (in log)
$\widehat{Closest\ mine\ is\ operating}$	0.060***	-0.163***	0.050***	0.144***	-0.157***	0.263***
	(0.002)	(0.007)	(0.002)	(0.007)	(0.023)	(0.007)
Distance to closest mine (in km)	-0.000***	0.020***	-0.001***	0.000	0.028***	-0.010***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000	-0.001***	-0.000***	0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.006***	0.021***	0.005***	-0.007***	0.035***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)
Number of mines in a ratio of 1 km	-0.046***	-0.011***	-0.057***	0.001	0.260***	0.044***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.024)	(0.004)
Number of mines in a ratio of 3 km	-0.005***	0.003***	-0.002***	-0.031***	-0.456***	-0.142***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.011)	(0.002)
Number of mines in a ratio of 5 km	0.002***	0.001**	0.001***	0.030***	0.014***	0.053***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)
Number of mines in a ratio of 10 km	0.001***	0.000	0.001***	-0.003***	0.048***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 25 km	0.000***	-0.002***	0.001***	-0.001***	-0.007***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	0.000***	0.001***	0.000***	-0.000***	0.003***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	0.001***	0.004***	0.000*	0.003***	-0.001	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Female	-0.005***	0.006*	-0.007***	0.029***	0.018**	-0.021***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.008)	(0.003)
Public school	0.418***	0.734***	0.418***	0.375***	-0.367***	0.356***
	(0.001)	(0.005)	(0.001)	(0.005)	(0.020)	(0.004)
Household income	0.041***	0.057***	0.033***	0.082***	-0.019***	0.046***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.006)	(0.001)
Father's years of education	0.008***	0.006***	0.008***	0.009***	-0.008***	0.014***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Mother's years of education	0.008***	0.006***	0.007***	0.015***	-0.016***	0.015***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Ethnicity	0.089***	0.061***	0.073***	-0.146***	-0.188***	0.283***
	(0.002)	(0.006)	(0.002)	(0.011)	(0.049)	(0.006)
Coed high school	-0.240***	0.020*	-0.267***	-0.189***		0.198***
	(0.001)	(0.011)	(0.002)	(0.010)		(0.006)
Urban high school	0.506***	0.475***	0.475***	0.544***	-0.082***	0.419***
	(0.001)	(0.005)	(0.001)	(0.003)	(0.020)	(0.004)
Academic degree	-0.194***	-0.180***	-0.196***	-0.148***	-0.189***	-0.235***
	(0.001)	(0.004)	(0.001)	(0.003)	(0.010)	(0.004)
School calendar from Jan to Dec	0.145***	-0.149***	0.124***	-0.084***		0.400***
	(0.002)	(0.008)	(0.002)	(0.009)		(0.011)
Size of the mine	-0.000***	-0.000***	0.000***	-0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	0.039***	0.022***	0.030***	-0.057***	-0.043***	0.096***
	(0.001)	(0.005)	(0.001)	(0.003)	(0.005)	(0.005)
School longitude	-0.016***	0.187***	-0.016***	0.060***	-0.013	0.112***
	(0.001)	(0.006)	(0.001)	(0.004)	(0.012)	(0.006)
Constant	0.000	7.960***	1.090***	0.561	0.000	0.000
	(0.000)	(0.925)	(0.195)	(0.673)	(0.000)	(0.000)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.178	0.262	0.170	0.241	0.300	0.311
IV F-Stat	3.309e+06	153559	2.781e+06	120348	25610	175608

Note: The table displays the coefficients of interest for the Size of cohort. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 3, with the regression incorporating time and departmental controls (not shown). First step for the IV approach in Table 2A. Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 8A. Saber 11 test score (IV)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Saber 11 score	Saber 11 score	Saber 11 score	Saber 11 score	Saber 11 score	Saber 11 score
Closest $\widehat{\text{mine is}}$ operating	0.011	0.671**	-0.309***	3.833***	-3.594***	-2.133***
	(0.055)	(0.306)	(0.061)	(0.268)	(0.721)	(0.252)
Distance to closest mine (in km)	-0.083***	0.030	-0.176***	-0.141***	-0.246***	0.134***
	(0.001)	(0.042)	(0.003)	(0.008)	(0.052)	(0.015)
$Distance\ to\ closest\ mine^2\ (in\ km)$	0.000***	-0.000	0.000***	0.000***	0.000***	-0.000***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.066***	0.090***	0.054***	-0.186***	-0.604***	0.113***
	(0.003)	(0.020)	(0.003)	(0.017)	(0.103)	(0.014)
Number of mines in a ratio of 1 km	-0.453***	-0.540***	-0.304***	0.058	1.268*	-1.095***
	(0.019)	(0.078)	(0.022)	(0.047)	(0.762)	(0.146)
Number of mines in a ratio of 3 km	0.073***	0.039	0.024***	0.174***	1.115***	0.702***
	(0.006)	(0.036)	(0.006)	(0.018)	(0.340)	(0.055)
Number of mines in a ratio of 5 km	0.040***	0.032	0.064***	-0.259***	0.052	-0.138***
	(0.003)	(0.022)	(0.003)	(0.015)	(0.102)	(0.027)
Number of mines in a ratio of 10 km	-0.055***	0.023***	-0.069***	-0.038***	-0.154***	-0.028***
	(0.001)	(0.007)	(0.001)	(0.005)	(0.029)	(0.005)
Number of mines in a ratio of 25 km	0.003***	-0.012***	0.003***	0.019***	-0.011*	0.013***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.006)	(0.002)
Number of mines in a ratio of 50 km	0.010***	0.004***	0.011***	0.017***	0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Year of birth	0.202***	0.521***	0.639***	0.558***	0.227***	0.333***
	(0.001)	(0.016)	(0.003)	(0.012)	(0.016)	(0.008)
Female	-4.158***	-3.565***	-4.182***	-3.526***	-3.942***	-3.881***
	(0.022)	(0.119)	(0.024)	(0.089)	(0.252)	(0.096)
Public school	-1.677***	-0.930***	-1.586***	-2.731***	11.300***	-6.346***
	(0.030)	(0.206)	(0.032)	(0.178)	(0.633)	(0.162)
Household income	4.566***	4.094***	4.537***	3.885***	3.000***	4.060***
	(0.013)	(0.070)	(0.014)	(0.053)	(0.175)	(0.053)
Father's years of education	0.685***	0.502***	0.675***	0.453***	0.703***	0.737***
	(0.005)	(0.029)	(0.006)	(0.020)	(0.061)	(0.024)
Mother's years of education	0.910***	0.679***	0.874***	0.788***	1.224***	0.948***
	(0.006)	(0.031)	(0.006)	(0.022)	(0.068)	(0.026)
Ethnicity	-1.113***	-2.812***	-0.609***	-7.624***	-11.339***	-0.064
	(0.054)	(0.233)	(0.059)	(0.407)	(1.541)	(0.207)
Coed high school	-14.190***	-8.984***	-13.805***	-6.963***		-15.504***
	(0.052)	(0.461)	(0.055)	(0.357)		(0.221)
Urban high school	1.705***	3.745***	1.890***	0.139	3.261***	0.449***
	(0.031)	(0.187)	(0.035)	(0.111)	(0.617)	(0.151)
Academic degree	-0.907***	-0.674***	-0.958***	-1.409***	-2.260***	1.016***
	(0.026)	(0.150)	(0.029)	(0.104)	(0.327)	(0.128)
School calendar from Jan to Dec	-4.413***	-7.822***	-3.947***	-2.757***		-8.993***
	(0.071)	(0.311)	(0.077)	(0.346)		(0.403)
Size of the mine	-0.000***	-0.000	-0.002***	-0.000***	-0.002***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	-1.093***	-1.972***	-1.553***	-3.903***	-2.430***	1.692***
	(0.027)	(0.194)	(0.032)	(0.113)	(0.170)	(0.173)
School longitude	4.841***	1.945***	2.440***	0.360**	4.608***	8.923***
	(0.032)	(0.253)	(0.043)	(0.145)	(0.367)	(0.217)
Constant	0.000	-833.749***	-1,042.201***	-1,016.934***	0.000	0.000
	(0.000)	(37.777)	(6.884)	(25.195)	(0.000)	(0.000)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.157	0.136	0.158	0.222	0.130	0.172
IV F-Stat	3.309e+06	153559	2.781e+06	120348	25610	175608

Note: The table displays the coefficients of interest for the Saber 11 test score. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 3, with the regression incorporating time and departmental controls (not shown). First step for the IV approach in Table 2A. Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 9A.Enrollment in Higher Education (IV)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education	Enrolled in higher education
$\widehat{\text{Closest mine is operating}}$						
	0.045***	0.004	0.036***	0.007	-0.115***	0.007
Distance to closest mine (in km)	(0.001)	(0.006)	(0.001)	(0.005)	(0.014)	(0.005)
	-0.002***	0.002**	0.000	-0.001***	0.004***	-0.001**
$Distance\ to\ closest\ mine^2\ (in\ km)$	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
	0.000***	-0.000***	-0.000***	0.000**	-0.000***	-0.000***
Mine operation time (in years)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.001***	-0.000	-0.001***	-0.001***	-0.014***	0.001***
Number of mines in a ratio of 1 km	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
	-0.006***	0.000	-0.003***	-0.002*	0.105***	-0.005**
Number of mines in a ratio of 3 km	(0.000)	(0.001)	(0.000)	(0.001)	(0.014)	(0.003)
	0.000	-0.001	-0.000**	0.000	0.031***	0.003**
Number of mines in a ratio of 5 km	(0.000)	(0.001)	(0.000)	(0.000)	(0.006)	(0.001)
	0.000***	0.000	0.001***	-0.000	0.018***	-0.000
Number of mines in a ratio of 10 km	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
	-0.000***	0.000	-0.000***	-0.000	-0.008***	-0.000**
Number of mines in a ratio of 25 km	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
	0.000**	-0.000***	-0.000	-0.000	-0.001***	0.000***
Number of mines in a ratio of 50 km	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.000***	0.000	0.000***	0.000***	-0.000***	-0.000***
Year of birth	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	0.005***	0.025***	0.026***	0.026***	0.008***	0.011***
Female	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.016***	-0.004*	-0.015***	-0.008***	-0.010**	-0.017***
Public school	(0.000)	(0.002)	(0.000)	(0.002)	(0.005)	(0.002)
	-0.046***	-0.059***	-0.054***	-0.060***	0.219***	-0.071***
Household income	(0.001)	(0.004)	(0.001)	(0.003)	(0.012)	(0.003)
	0.039***	0.048***	0.038***	0.036***	0.042***	0.034***
Father's years of education	(0.000)	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)
	0.003***	0.002***	0.003***	0.003***	0.002**	0.003***
Mother's years of education	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
	0.007***	0.005***	0.005***	0.005***	0.007***	0.006***
Ethnicity	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
	-0.020***	-0.034***	-0.010***	-0.079***	-0.394***	-0.003
Coed high school	(0.001)	(0.004)	(0.001)	(0.008)	(0.029)	(0.004)
	-0.115***	-0.129***	-0.098***	-0.023***		-0.107***
Urban high school	(0.001)	(0.009)	(0.001)	(0.007)		(0.004)
	0.036***	0.034***	0.034***	0.023***	0.137***	0.053***
Academic degree	(0.001)	(0.003)	(0.001)	(0.002)	(0.012)	(0.003)
	-0.011***	0.006**	-0.009***	-0.018***	-0.023***	0.007***
School calendar from Jan to Dec	(0.000)	(0.003)	(0.001)	(0.002)	(0.006)	(0.002)
	-0.100***	-0.112***	-0.100***	-0.090***		-0.191***
Size of the mine	(0.001)	(0.006)	(0.001)	(0.007)		(0.007)
	0.000***	-0.000***	-0.000***	0.000***	0.000	0.000***
School latitude	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.003***	-0.004	0.001**	-0.013***	-0.028***	-0.005
School longitude	(0.001)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)
	0.121***	0.010**	0.007***	-0.020***	0.189***	0.293***
Constant	(0.001)	(0.005)	(0.001)	(0.003)	(0.007)	(0.004)
	0.000	-48.710***	-50.411***	-51.312***	0.000	0.000
Observations	(0.000)	(0.699)	(0.124)	(0.479)	(0.000)	(0.000)
R^2	5,710,986	190,625	4,843,896	321,301	42,258	312,906
IV F-Stat	0.046	0.081	0.079	0.084	0.063	0.060

Note: The table displays the coefficients of interest for the probability of enrollment in higher education. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 3, with the regression incorporating time and departmental controls (not shown). First step for the IV approach in Table 2A. Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 10A. Enrollment in labor market (IV)

	A. Full sample	B. By extracted product				
		Co-As	Construction	Gold	Metals	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market	Enrolled in formal labor market
$\widehat{\text{Closest mine is operating}}$	-0.002***	-0.010***	0.011***	0.010***	0.046***	-0.016***
	(0.001)	(0.003)	(0.001)	(0.003)	(0.008)	(0.002)
Distance to closest mine (in km)	0.001***	-0.001***	0.000***	0.000	-0.005***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
$Distance\ to\ closest\ mine^2\ (in\ km)$	-0.000***	0.000**	-0.000***	-0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mine operation time (in years)	0.000***	0.001***	-0.000***	-0.000	0.010***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 1 km	0.003***	-0.001	0.002***	0.001*	-0.031***	0.002*
	(0.000)	(0.001)	(0.000)	(0.000)	(0.008)	(0.001)
Number of mines in a ratio of 3 km	-0.001***	-0.001***	-0.001***	-0.000	-0.037***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.001)
Number of mines in a ratio of 5 km	0.000***	0.001***	-0.000	-0.000	-0.006***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Number of mines in a ratio of 10 km	0.000***	-0.000***	0.000***	0.000	0.005***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 25 km	0.000	0.000	0.000***	0.000	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of mines in a ratio of 50 km	0.000***	-0.000	-0.000**	-0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth	-0.003***	-0.019***	-0.021***	-0.018***	-0.005***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.021***	-0.030***	-0.022***	-0.027***	-0.035***	-0.020***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)
Public school	-0.007***	0.006***	-0.001***	0.003*	-0.100***	-0.012***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.007)	(0.002)
Household income	-0.003***	-0.003***	-0.002***	-0.004***	-0.007***	-0.001**
	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Father's years of education	-0.000***	0.000	0.000	-0.000	-0.001	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Mother's years of education	-0.002***	0.000	-0.000***	-0.000	0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Ethnicity	0.006***	0.001	0.000	-0.003	0.254***	-0.000
	(0.001)	(0.002)	(0.001)	(0.004)	(0.017)	(0.002)
Coed high school	0.033***	0.011**	0.021***	-0.001		0.022***
	(0.001)	(0.004)	(0.001)	(0.004)		(0.002)
Urban high school	-0.005***	-0.001	-0.001***	0.003**	-0.059***	-0.035***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.007)	(0.001)
Academic degree	0.005***	0.007***	0.004***	0.003***	0.003	-0.016***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.004)	(0.001)
School calendar from Jan to Dec	-0.012***	0.003	-0.010***	-0.027***		0.022***
	(0.001)	(0.003)	(0.001)	(0.004)		(0.004)
Size of the mine	-0.000***	0.000	-0.000	-0.000	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
School latitude	0.003***	0.000	-0.003***	-0.003**	0.016***	0.011***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.002)	(0.002)
School longitude	-0.093***	0.008***	0.003***	-0.008***	-0.112***	-0.194***
	(0.000)	(0.002)	(0.000)	(0.001)	(0.004)	(0.002)
Constant	0.000	38.306***	41.555***	35.243***	0.000	0.000
	(0.000)	(0.366)	(0.066)	(0.260)	(0.000)	(0.000)
Observations	5,710,986	190,625	4,843,896	321,301	42,258	312,906
R^2	0.047	0.104	0.120	0.105	0.069	0.052
IV F-Stat	3.309e+06	153559	2.781e+06	120348	25610	175608

Note: The table displays the coefficients of interest for the probability of enrollment in the labor market. Panel A presents the results for the complete dataset, while Panel B divides the results by the type of extracted product. The estimations follow the specifications outlined in Equation 3, with the regression incorporating time and departmental controls (not shown). First step for the IV approach in Table 2A. Co-As represents Coal and Asbestos. Robust standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1.