



Local incentives and national tax evasion: The response of illegal mining to a tax reform in Colombia[☆]

Santiago Saavedra^a, Mauricio Romero^{b,*}

^a Universidad del Rosario, Department of Economics, Colombia

^b ITAM, Centro de Investigación Económica, Mexico

ARTICLE INFO

JEL classification:

H26

O13

O17

Keywords:

Tax evasion

Illegal mining

Satellite images

Colombia

ABSTRACT

Achieving a fair distribution of resources is one of the goals of fiscal policy. To this end, governments often transfer tax resources from richer to more marginalized areas. In the context of mining in Colombia, we study whether lower transfers to the locality where the taxed economic activity takes place dampen local authorities' incentives to curb tax evasion. Using machine learning predictions on satellite imagery to identify mines allows us to overcome the challenge of measuring evasion. Employing difference-in-differences strategies, we find that reducing the share of revenue transferred back to mining municipalities leads to an increase in illegal mining. This result highlights the difficulties inherent in adequately redistributing tax revenues.

1. Introduction

Achieving a fair distribution of resources is one of the goals of fiscal policy. To this end, governments often transfer tax resources from richer to more marginalized areas. Yet, a smaller share of revenue transferred back to the locality where the taxed economic activity takes place could dampen local authorities' incentives to curb tax evasion (Banerjee and Hanna, 2012). This is a classic agency problem in tax collection and evasion enforcement between central and lower levels of government. We study this dimension of tax evasion in the context of illegal mining in Colombia. We find that reducing the share of mining royalties transferred back to the mining municipalities leads to an increase in illegal mining.

One challenge to studying tax evasion (or any illegal activity) is measuring its extent (Banerjee and Hanna, 2012; Slemrod and Weber, 2012; Fisman and Wei, 2004). To overcome this issue, we construct a novel dataset using machine learning predictions on satellite imagery to detect mining activity. We assess the legality of identified mining activity using georeferenced mining permits issued by the central government.¹ The specific mineral mined in each location is approximated, based on research conducted

[☆] We are grateful to Prashant Bharadwaj, Arun Chandrasekhar, Pascaline Dupas, Larry Goulder, and Melanie Morten for their invaluable guidance. This paper benefited from insightful comments on the part of Juan Castillo, Thomas Ginn, Marcel Fafchamps, Mounu Prem, Mark Rosenzweig, Jeff Weaver, anonymous referees, and seminar participants at LACEA, NEUDC, Stanford University, UC San Diego, Banco de la República, Universidad de los Andes, Universidad del Rosario, PacDev, and Ridge. Francisco Barrera, Muhammad Islam, Fabian Latorre, Sergio Perilla, and Michael Wang provided excellent assistance with the construction of the data. This study was carried out with support from the Latin American and Caribbean Environmental Economics Program (LACEEP), Costa Rica and Policy Design and Evaluation Lab (PDEL), United States. Replication code and data for this article can be found at <https://doi.org/10.7910/DVN/NMLOMG>. Saavedra acknowledges support of the Stanford Institute for Economic Policy Research, United States through the Kohlhaugen Fellowship Fund. Romero acknowledges financial support from the Asociación Mexicana de Cultura, Mexico.

* Corresponding author.

E-mail addresses: santiago.saavedrap@urosario.edu.co (S. Saavedra), mtromero@itam.mx (M. Romero).

¹ We measure illegal mining following the Colombian government's definition: "mining activity without a mining permit registered with the National Mining Registry" (Ministerio de Minas y Energía, 2003, p. 108). According to the 2010 Mining Census, among mines without a permit, 85% evade royalty taxes.

by the National Mining Agency on the potential subsoil resources across the country (Agencia Nacional de Minería, 2013). Our measurement of illegal mining allows us to study the effect of a reform in Colombia that sharply reduced the share of mining tax revenue transferred back to the municipality where the mine is located.² Since mining tax enforcement is relegated to the local government — a common practice in countries with fiscal federalism (OECD, 2018, 2019a,b) — this reform reduced the incentives of local authorities to monitor miners' compliance with the central government's regulations.³

Traditional tax evasion models with exogenous enforcement, which do not take into account where tax revenue is spent, will predict a null effect of the reform (Slemrod, 2019). We consequently present a simple theoretical framework — in which a miner decides whether or not to operate legally — to illustrate how the reform affects the incentives of local authorities to overlook illegal mining. We assume that the local authority observes mining activity in their municipality.⁴ Thus, to operate illegally, the miner must pay a bribe, which is determined by bargaining with the local authority. While the reform does not affect the cost of operating legally, it does lower the amount of tax revenue received by the municipality from legal mines. As legal mining has a lower payout for the local authority after the reform, the incentives to enforce regulations diminish.

The model yields four predictions.⁵ First, mines are more likely to operate illegally after the reform. Second, the effect is larger for minerals with a higher royalty tax rate. Third, the impact of the reform is greater in areas where the central government's presence is weak. Finally, the reform should not affect illegal mining in municipalities where illegal armed groups are the *de facto* local authorities.

We employ two different identification strategies to test these predictions. While neither strategy is flawless, they both suggest that the reform caused an increase in illegal mining. First, a difference-in-differences strategy comparing minerals across time within the same Colombian municipality allows testing whether illegal mining increased more for those minerals with a higher tax rate. For example, since gold has a higher royalty rate than coal, we expect a larger increase in illegal mining in gold-mining areas compared to coal-mining areas, even within the same municipality. We control for municipality-time fixed effects that capture time-varying confounding events, including a new mayor taking office, the peace process with guerrillas, and changes in the mining permit request system in Colombia.⁶ We also control for municipality-mineral fixed effects, which absorb differences in the mineral-specific production function in each municipality that are constant over time.

We find that for every percentage point of the royalty tax rate, there was a 1.18 percentage point increase in the area being mined illegally after the reform. This increase in illegal mining translates into approximately 218 million USD of foregone revenue per year, equivalent to 37% of the 594 million USD in mining royalties in 2015.

To further test whether mines are more likely to operate illegally after the reform, we again use a difference-in-differences strategy, this time comparing illegal mining in Colombia and Peru, before and after the reform. As a share of the total mined area, illegal mining increased in Colombia by 4.21 percentage points after the reform relative to the change observed for Peru. While the treatment estimate may be confounded by other changes that could affect illegal mining in either of the two countries, we specifically chose Peru as our control to minimize this risk (Section 5.1 provides more details). In addition, this result is robust to different specifications and various ways of defining the control group. While no single piece of evidence addresses all identification concerns, taken together, our findings illustrate that the reform increased illegal mining in Colombia.

Finally, we use cross-sectional variation in municipality characteristics to identify the heterogeneous effect of the reform on municipalities where the central government's presence is weak (prediction 3) or where control is held by an illegal armed group (prediction 4). As predicted by the model, the impact of the reform is greater in municipalities where the central government's presence is weak. Likewise, the effect is smaller in municipalities run by illegal armed groups (i.e., not effectively governed by local authorities).

This is the first paper to quantify the response of tax evasion to the share of tax revenue allocated to local spending to the best of our knowledge.⁷ Our study joins an extensive literature on the determinants of tax evasion (see Slemrod, 2019 for a review). Notably, Khan et al. (2015) present experimental evidence that performance pay for tax collectors increases both tax revenue and reported bribes. While local authorities are not direct tax collectors in our context, we find a similar effect when incentives for tax collection are reduced.

The results herein also build on the public finance literature on how revenues should be collected and distributed among different levels of government (Gadenne and Singhal, 2014). Previous papers have shown that: (i) politicians value spending on projects for citizens in their home areas twice as much as general government spending (Hoffman et al., 2017); and (ii) when local governments

² The central government distributes the royalty revenue that is not transferred to mining municipalities among all the country's municipalities according to socioeconomic indicators.

³ As discussed below, the reform did not alter municipalities' capacity to monitor enforcement since it is the central government that funds local police forces and the national army.

⁴ This is supported empirically by a survey of 18 local authorities; all confirmed they were aware of the presence of illegal mining within their jurisdictions (Fedesarrollo, 2014).

⁵ Similar to Slemrod and Yitzhaki (2002), the effect of the reform on evasion depends on functional form assumptions. If the probability of detection by the central government and the penalty are constant irrespective of mine size, the reform would have no effect. However, larger mines are easier to observe, and Colombian law allows confiscation/destruction of machinery of illegal mines. Therefore, we assume both that bigger mines have a higher probability of detection and suffer a greater penalty if caught (i.e., they have more machinery).

⁶ This identification strategy is very similar to that employed by Sánchez de la Sierra (2020) in their study of mineral taxation in the Democratic Republic of the Congo. However, we exploit within-municipality variation rather than that across municipalities.

⁷ While Cai and Treisman (2004) provide examples of specific cases where local governments helped firms evade federal taxes and regulations; we provide causal evidence using quasi-experimental variation.

collect tax revenue directly (as opposed to receiving grants from the central government) they spend more efficiently (Gadenne, 2017). Our paper complements these two findings by showing that raising and spending taxes locally has the advantage of reducing evasion. This result is similar in spirit to that of Zhuravskaya (2000), who demonstrates that cities in Russia do not invest much effort in raising tax revenue, given that most of what they raise is transferred to the central government.⁸

Finally, this paper contributes to the small but burgeoning literature on the political economy of natural resource management. Like Burgess et al. (2012) and Lipscomb and Mobarak (2016), we study a national interest resource, the monitoring of which depends on local authorities. Both of these studies show that the political incentives of local government officials increase deforestation and water pollution. We complement their findings by demonstrating that fiscal incentives are also an important determinant of illegal mining, one of the drivers of Amazonian deforestation (Swenson et al., 2011; Asner et al., 2013). In a closely related paper, Shapiro and van den Eynde (in press) assess a royalty reform in India that increased taxes for iron ore, and find that legal mines increase evasion. Our setting differs in that it was not taxes, but the revenue allocation that changed.

Methodologically, we employ machine learning to both construct the dependent variable and to estimate causal effects. In using applications of machine learning techniques for causal inference (Belloni et al., 2014; Athey and Imbens, 2016), we join the growing body of literature that utilizes satellite observations to study economic outcomes (e.g., Foster et al., 2009; Jayachandran, 2009; Henderson et al., 2012; Guiteras et al., 2015; Faber and Gaubert, 2019). In contrast, previous studies on illegal mining use static measures in their analyses (Idrobo et al., 2014; Romero and Saavedra, 2015). Our municipality panel dataset on illegal mining is thus a contribution in and of itself. This is similarly true of the algorithm that creates the dataset, which could be employed to create comparable data for other countries.

2. Mining context and details of the reform

2.1. Mining in Colombia

The mining and hydrocarbon industry generated 8%–11% of Colombia's Gross Domestic Product around the time of the reform.⁹ Although mineral mining represents only 20% of royalty tax revenue, it has a large footprint — large enough that its environmental impacts can be tracked from space (Asner et al., 2013). Within mineral mining, 77% of the royalties come from coal, 12% from nickel, 10% from precious metals (e.g., gold, silver, and platinum), and the remaining fraction from salt, emeralds, and construction materials.

According to Colombia's constitution (article 332), the central government owns subsoil resources, including minerals. This ownership is different from other countries, such as the United States, where the owner of the land holds the rights to the mineral resources it contains.¹⁰ Colombia's central government allocates mining permits and sets royalty taxes for mineral extraction. The permit holder pays a fee equal to a daily minimum wage per hectare per year.¹¹ In addition, permit holders must pay royalties based on the gross value and type of minerals extracted.¹² Royalties vary across minerals and depend on the quantity extracted. For example, the rate for construction materials is 1%, for platinum and coal is 5%, and the rate for alluvial gold is 6%. Table A.1 in the Appendix provides details of the royalty rate for each mineral.

Holding a mining permit in the neighboring country of Peru, broadly speaking, has the same implications as in Colombia. All natural resources belong to the central government in both countries, and mining permit holders must follow environmental regulations and pay associated royalties. Ministerio de Minas y Energía (2016) provides more details on the Peruvian regulation. In Section 5.1 we discuss our reasons for choosing Peru as the control country for one of the identification strategies.

2.2. The reform

Before 2012, a municipality in Colombia received around 55% of the royalties paid by mines operating in its territory. Legislative Act 05 of 2011 changed the allocation formula, such that only 10% of the royalties are now transferred directly to the mining municipality. Another 40% of royalties revenue is earmarked for regional funds, and the rest must be saved. Regional funds are distributed according to population, poverty, and unemployment rates; thus, the net impact in the budget of each municipality varies according to these characteristics.¹³

The Colombian Congress approved the reform in July 2011, and it went into effect in January 2012. The reform resulted in an average increase of 4% in the royalty transfers received by municipalities relative to their total budget (Table 1, Panel A). While this

⁸ In principle, tax morale could also explain our results, i.e., citizens evade more because the revenue is not spent on their municipality (Falkinger et al., 1988; Cullen et al., in press). However, in our context, less than 20% of mine owners are from the Colombian municipality where their mine is located (Fletcher and Saavedra, 2019), making it unlikely that tax morale explains our results.

⁹ This is common across developing countries. Total natural resources rents accounted for 11.7% of GDP across low-income countries (World Bank, 2017).

¹⁰ However, this is common in many countries. For example, besides Colombia and Peru, in Brazil, Ecuador, and Guyana, all mineral resources are owned by the state, including the minerals on and under indigenous lands. See https://publications.wri.org/undermining_rights/data-and-findings.

¹¹ If the permit area is between 2,000 and 5,000 hectares, the permit holder pays two times the legal minimum wage per hectare. While holders of permits for areas larger than 5,000 hectares pay three times the minimum wage per hectare (Agencia Nacional de Minería, 2013).

¹² The price used to calculate the gross value is the average monthly price on the London Metal Exchange. Colombia is considered a price taker in all mineral markets, given the scale of its production (Fedesarrollo, 2014). For example, Colombia accounts for ~1% of the world's coal and gold production.

¹³ The reform requires that 10% of royalties be allocated to a science, technology, and innovation fund; 10% must be allocated to underfunded pensions; and up to 30% are placed in a savings and stabilization fund.

Table 1
Summary statistics for municipalities used in the analysis.

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Median	Std. Dev.	Min	Max	N
Panel A: Colombian Municipalities budget change after the reform						
% budget change	4.11	7.51	11.66	-62.5	52.3	927
% budget loss if loss	-16.91	-12.52	15.02	-62.5	-0.0	150
% budget win if won	8.17	8.24	4.13	0.0	52.3	777
Panel B: Characteristics Colombian Mining Municipalities						
Population	23,471.53	12,796.00	36,282.56	861.0	377,693.0	927
Area (km ²)	641.72	269.28	1,328.02	15.4	17,266.0	927
Area mining titles (km ²)	11.60	0.76	54.84	0.0	1,295.4	927
Panel C: Characteristics Peruvian Municipalities						
Population	14,864.54	4,442.00	44,595.66	191.0	860,107	1,787
Area (km ²)	627.27	207.44	1,624.63	1.9	22,184	1,806
Area mining titles (km ²)	20.21	0.81	71.11	0.0	2,020	1,806
Panel D: Machine Learning Training Data						
On 2010 Mining Census	0.55	1.00	0.50	0.0	1.0	927
% illegal area (Census)	36.49	8.84	42.34	0.0	100.0	505
% open pit mines (Census)	78.19	100.00	35.09	0.0	100.0	505
% illegal area open pit (Census)	40.09	13.39	43.73	0.0	100.0	472
% illegal area (UNODC)	80.79	94.79	23.32	0.0	100.0	913
% illegal area (UNODC) Censed	80.29	92.02	22.74	0.3	100.0	500

Notes: In Panels A, B and D, an observation is a Colombian municipality. Although there are 1,122 municipalities in Colombia, we include only those with minerals in the subsoil: 927 municipalities. In Panel A, % budget change refers to the change in royalties received by the municipality 2013–2014 compared to 2010–2011, as percentage of the municipality budget 2010–2011. The next two rows separate by those with negative change (“loss”) and positive change (“won”). In Panel B, the population is from the 2005 Population Census; area mining titles calculated with National Mining Agency data. In Panel D, the rows related to the Census only have information for municipalities that were covered by the 2010 Mining Census. United Nations Office on Drug and Crime (UNODC), refers to the map of illegal gold mines produced by this entity with manual validation. Source: CEDE panel data, 2010 Mining Census, and UNODC. In Panel C, an observation is a Peruvian municipality. There are 1,787 municipalities. Population data is from 2004. Source: National Statistics Bureau (INEI).

masks heterogeneity by how the redistribution process affected each municipality — as the reform allocated some funds according to population and poverty and unemployment rates — the variation generated by the redistribution process is relatively small. That is, the marginal change in revenue generated by the redistribution process is an order of magnitude smaller than the change from 55% to 10% of direct transfers. Since most royalties come from oil and gas, mostly municipalities with oil or gas fields lost revenue from the reform. Most mining municipalities ended with larger royalty transfers after the reform (except if they had oil or gas). However, the marginal revenue from a given mine decreased considerably after the reform.

The objectives of the reform were to reduce poverty and regional inequality, save part of the expected increase in mining revenue, and improve the management of royalty resources.¹⁴ Illegal mining was not mentioned as a motivation for the reform, nor were the impacts of the reform on illegal mining contemplated.¹⁵

The reform did not change the permit fees paid to operate legally, nor did it change the royalty rate paid by mines. Likewise, the reform did not alter the broad spending categories for which municipalities can use royalty revenue. Local governments are responsible for the provision of education, health, drinking water, and sanitation services.¹⁶ Municipalities must use over 90% of royalty transfers to meet basic needs related to these services until certain goals are achieved.¹⁷

2.3. Illegal mining

Illegal mining is widespread in Colombia, according to various sources. The central government conducted the 2010 mining census, including legal and illegal mines (without a mining permit). According to the census, 36% of the reported mined area was not under a mining permit.¹⁸ The United Nations Office on Drug and Crime (UNODC) manually inspected satellite images from

¹⁴ See <https://www.sgr.gov.co/LinkClick.aspx?fileticket=bsf8qrvgVog%3D&tabid=181> for the motivation of the reform. Martinez (2019) provides evidence of the mismanagement of royalty resources.

¹⁵ Congress approved the reform six months before implementation. Thus, we cannot rule out some anticipation by local governments of its effects. However, we argue that the timing of the reform is exogenous to the evolution of illegal mining. Appendix B.1 provides a detailed timeline of the reform.

¹⁶ Roughly 90% of the revenue for local authorities comes from transfers — including royalties — from the central government. The main source of revenue raised by local authorities — over which there is almost complete discretion in their use — comes from property taxes (Martinez, 2019).

¹⁷ The goals are: infant mortality under 1%, universal health coverage, net primary enrollment over 90%, access to safe drinking water over 70%, and access to sewage systems over 70%. After these goals are met, municipalities must allocate 75% of the revenue to these basic services and the rest can be used for environmental conservation and restoration, other education and health projects, or other public works (Ministerio de Minas y Energía, 1995; de Planeación, 2007).

¹⁸ Even if a person has ownership of the land, they need a permit to mine because the central government owns the subsoil minerals.

2014 and found that 81% of the area mined for gold was illegally mined (UNODC, 2016). The UNODC (2016) estimate of 81% of area illegally mined is comparable to the 84% estimate we generate using machine learning with satellite images (Section 4.2 provides more details on our estimates). Both the UNODC (2016) and our estimates of illegal mining are higher than the census estimates. This fact is consistent with some illegal mine operators refusing to answer the census or underreporting their mining area. We cannot use the 2010 census directly in our empirical strategy since it is not a panel.

There have been three attempts to legalize illegal mines in Colombia, all with little success.¹⁹ Government incentives for illegal mines' legalization have also been accompanied by stricter enforcement measures: in late 2012, the Andean Community, of which Colombia and Peru are members, issued a decree authorizing the destruction of all machinery used in mines that do not have a registered permit.²⁰

Local authorities in Colombia are responsible for enforcing mining laws in their areas. This responsibility includes suspending any mining activity carried out without a permit, according to Law 685 of 2001. If the suspension notification is ineffective, local authorities must inform national law enforcement agencies (i.e., the police and the army) that they should proceed with the confiscation/destruction of machinery. In addition, they must confiscate any minerals without a certificate of origin from a legal mine. However, there are several documented cases in which local authorities are often aware of illegal mines and receive bribes from illegal miners (e.g., see *Semana*, 2013; *Pardo*, 2013; *Solarte*, 2017; *El Espectador*, 2017). The Inspector General's Office (*Procuraduría General de la Nación*) is in charge of investigating local authorities who do not comply with the law. From 2009 to 2017, 37 disciplinary procedures were initiated against local authorities (mayors or state governors) for allowing (implicitly or explicitly) illegal mining in their jurisdiction. The number of disciplinary procedures per year is higher after the reform, suggesting local authorities became more likely to disregard their duties concerning illegal mining (see *Fig. A.1*).²¹

The central government funds both local police forces and the national army — but the mayor is in command of the police force of each municipality.²² Thus, the reform did not alter the funding mechanism for enforcement agencies. With enforcement unaffected, the reform still can affect illegal mining because local authorities are more willing to accept bribes. This channel is what we proceed to illustrate with the theoretical framework.

3. Theoretical framework

We present a simple framework for understanding a miner's decision to operate legally or illegally based on bargaining with the local authority. The decision depends on the taxes and fees if operating legally, and on the penalty and probability of detecting illegal mines by the central government. We derive four theoretical predictions based on reasonable assumptions.

Intuitively, local authorities are willing to “overlook” illegal mining if the miners pay a bribe proportional to the cost of operating legally. Since the reform reduces the share of royalties transferred to the municipality from legal mining, more miners can pay the amount to local authorities to “overlook” illegal operations (**Prediction 1**). As the reform reduction is larger for minerals with higher royalty rates, we expect a higher increase in illegal mining for minerals with higher royalty rates (**Prediction 2**). In municipalities with lower national oversight, the effect of the reform is larger because larger mines can operate illegally without being detected (**Prediction 3**). Finally, if the local government has little power to enforce the law when mines operate illegally, the reform does not change the incentives of the different parties. Thus, the effect of the reform is smaller in places where illegal armed groups hold the de-facto power (**Prediction 4**).

This simple model has some limitations. First, it does not consider the location decision of the miner. While mineral resources are fixed in the subsoil, a miner could move its capital to a neighboring municipality, where conditions are more favorable à la *Burgess et al.* (2012). The capital reallocation across municipalities could amplify or reduce the effect of the reform depending on the fixed cost of capital relocation.²³ In the empirical section, we assess if the result is different in municipalities close to the border, where there could be more capital reallocation. Second, we abstract from income effects when a local authority receives bribes from many miners. Finally, we model the decision to get a permit and pay royalties as a single choice. Still, some legal mines may evade a percentage of production taxes (i.e., intensive margin evasion). While we mainly focus on the extensive margin evasion (i.e., having a permit), we assess the reform's effect on the quantity reported by legal mines in the empirical section. In addition, since production is difficult to monitor, any bargaining between the miner and the local authority is likely over the mine's area, which the local authorities can observe.

¹⁹ The 2001 Mining Code's legalization requirements were stringent, and of the 2,845 legalization requests received, only 23 were approved. Similarly, the Mining Code of 2010 generated 700 requests, but only approved one legalization. Finally, a pilot legalization program that started with 150 mining operations in 2012 only had 25 still in the process after three years, and none had met all the requirements (*The Global Initiative Against Transnational Organized Crime*, 2016).

²⁰ Before the 2012 decree was issued, the machinery was supposed to be confiscated, which was challenging to implement in remote regions.

²¹ An increase in disciplinary procedures could also reflect a response by the central government to counteract lower monitoring effort from the local authorities. However, since the Inspector General's Office is independent of the central government, the increase likely reflects a change in the local authorities' behavior.

²² See Article 315 of Colombia's constitution.

²³ Capital could move across countries or minerals, because of the reform. Suppose only capital from legal mines moves (from “control” to “treated” locations/minerals). In that case, this will go against the model's predictions, as it would increase the proportion of illegal mines in “control” locations. The opposite is true if only illegal capital moves.

3.1. Setup

The “surplus” of illegal mining for a mine with given capital K is the difference between the payoffs for the miner and the local authority when the operation is legal or illegal:²⁴

$$S(K) = \underbrace{T + \alpha pq(K)}_{\text{Legal mining fees}} - \underbrace{\beta \alpha pq(K)}_{\text{Foregone revenue}} - \underbrace{Pr(K)\Theta(K)}_{\text{Expected punishment}} \tag{1}$$

where T are the permit fees; α is the royalty tax rate the firm pays; p is the international price of the mineral; and $q(K)$ is the quantity extracted. The function $q(\cdot)$ is assumed to be increasing and concave.

β is the share of royalties allocated to the mining municipality — the parameter affected by the reform studied in this paper. $Pr(K)$ is the probability that the central government detects the illegal mine independently from the local authority. We assume $Pr(\cdot)$ is increasing, because empirically larger mines are easier to observe. $\Theta(K)$ is the penalty for the miner and the local authority if the mine is detected. Given that Colombian law dictates confiscation/destruction of machinery, the penalty for larger mines is larger (as they have more capital). That is, we also assume $\Theta(\cdot)$ is increasing.

Any firm with capital K such that $S(K) \geq 0$ will pay the bribe and operate illegally.²⁵ We denote by K^* the value of capital such that $S(K^*) = 0$. Any firm with $K > K^*$, and thus $S(K) < 0$, will operate legally (assuming returns to capital do not increase faster than expected punishment of capital destruction if caught operating illegally). The share of the mined area mined illegally (y) is thus:

$$y = \frac{\int_0^{K^*} Area(K) dK}{\int_0^{K^{max}} Area(K) dK} \tag{2}$$

3.2. The effect of the reform on illegal mining

As mentioned above, the reform did not change the fees paid to operate legally, nor did it change the royalty rate (α) paid by each mining firm. That is, the “Legal mining fees” part of $S(K)$ is unaffected by the reform. Regarding the “Foregone revenue” term, a reduction in the share of royalties transferred back to the mining municipality (β), reduces the payout from legal mining to the local authority. This translates into a higher illegal mining surplus for all mine sizes and a higher K^* . Thus, in Eq. (2) the numerator is larger. This analysis leads to the following prediction:

Prediction 1. *The reform increases the share of the mined area mined illegally.*

This decision model applies not only to a miner choosing to open a new mine. It also applies to a miner choosing to operate legally, if currently, he is not. In the empirical section, we will test both whether the cumulative and new area of illegal mines change.

Fig. 1 shows how the effect of the reform on illegal mining depends on the assumptions on the probability of detection ($Pr(K)$) and punishment ($\Theta(K)$). If the punishment and probability of detecting an illegal mine are constant, illegal mining is independent of the share of taxes for the local municipality. But only assuming increasing probability of detection, a smaller β increases illegal mining, since the surplus of illegal mining is larger for every level of capital.

In Appendix C.1, we show how different valuations of the municipality’s budget by the local authority affect this result. While we always assume that the valuation is positive and increasing, the results vary depending on whether the valuation is linear or convex. If the valuation is linear, the effect of the reform on illegal mining is the same for all municipalities regardless of whether they experience a net loss or win with the reform (i.e., there is no income effect). However, when the valuation is convex, there should be a larger increase in illegal mining in municipalities where the budget is negatively affected by the reform. The valuation is convex when local authorities capture an increasing share of the budget (Brollo et al., 2013), or there are increasing returns to scale from royalty income.²⁶ Fig. C.1 in the Appendix illustrates the effect of the reform for different shapes of the valuation of the budget.

3.3. Heterogeneous effects of the reform

Consider the change in the surplus of illegal mining (Eq. (1)), before and after the reform:

$$\Delta S(K) = \alpha_i p_i q_i (\beta_0 - \beta_1) \tag{3}$$

The subscript i highlights that the value of the parameters is different for each mineral i . Since the royalty rate (α_i) multiplies the change in the royalties share for the mining municipality, the change in the area illegally mined should be larger for minerals with a higher royalty rate.

²⁴ Appendix C.1 provides a detailed derivation of this expression.

²⁵ The model assumes K is exogenous and follows a uniform distribution. Firms are unable to merge.

²⁶ An illustration of convex valuation is the case of discrete investments. For example, with a small budget, only a vaccination campaign could be funded, while with a large budget, a hospital could be built, which is politically more visible. The median Colombian municipality spent 53 % of the revenue on “lumpy” projects such as the construction of a hospital or a bridge.

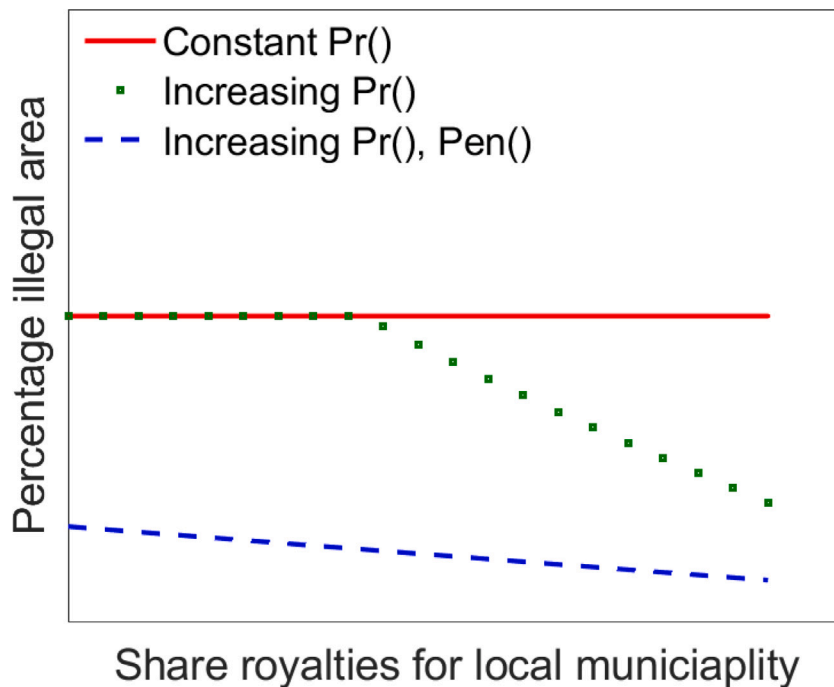


Fig. 1. Share of royalties for the local municipality and illegal mining. *Notes:* The graph shows the relationship between the share of royalty taxes allocated for the local municipality and illegal mining. The x -axis is the share of royalties allocated for the local municipality, β on the theoretical framework of Section 3. A point further to the right indicates a higher β . The y -axis is the percentage of the mined area mined illegally: a higher point represents a higher share of illegal mining. The red solid line plots the case of constant probability of detection $Pr(\cdot)$ and constant penalty $\theta(\cdot)$ in Eq. (1), irrespective of mine size. The green dots plot the case of increasing probability of detection for larger mines. The blue dashed line presents the case of increasing penalty and increasing probability of detection for larger mines. Larger mines are easier to observe and given that Colombian law allows confiscation/destruction of machinery, the penalty for larger mines is larger. Consequently, the blue dashed line represents our base assumptions.

Prediction 2. *After the reform, the increase in the share of area illegally mined is larger for minerals with a higher royalty rate.*

Consider a municipality where the probability of detecting illegal mines is small because of the central government's weak presence ($Pr_{WP}(\cdot) < Pr(\cdot)$). With a smaller probability of detection, the surplus of illegal mining is higher in these municipalities for any mine size ($S_{WP}(K) \geq S(K), \forall K$). When the reform reduces the share of royalties for mining municipalities, the surplus of illegal mining is positive for larger mines. Suppose the capital distribution is smooth at the threshold. In that case, we expect the reform to have a larger effect on the share of the area mined illegally in municipalities with a low probability of detection.²⁷

Prediction 3. *After the reform, the increase in the share of the area mined illegally is larger in municipalities with a lower probability of illegal mines being detected.*

So far, we have assumed that the local authority has some bargaining power with the miner. Thus, the royalties the local authority receives enter into the surplus calculation. But if the miner has full bargaining power and can ignore the local authority completely, the change in β does not change his legality decision. This case can apply in Colombian municipalities in areas where the local power is *de facto* held by illegal armed groups.²⁸ In a less extreme case, the effect of the reform is smaller.

Prediction 4. *There is a smaller effect of the reform if the local authority has little bargaining power with the miner.*

4. Data

We rely on four main sources of data for our analysis. The first is the panel of illegal mining by municipality that we constructed using machine learning; we provide details in the next subsection. The second database is from Colombia's central government

²⁷ In the extreme case in which armed groups have total control, and the central government is unable to destroy illegal mining machinery, all mines should be illegal ($S_{WP}(K) \geq 0, \forall K$), miners pay no royalties, and the reform should have no effect. However, this extreme case is not what we observe in the data: There are legal mines and royalties paid in municipalities with armed groups.

²⁸ We abstract from an endogenous response of armed groups. There is no evidence of armed group relocation in response to the budget changes of the reform (Table A.2 in the Appendix provides more details).

mineral information system, SIMCO, on reported production.²⁹ We also use the municipality panel from the Center for Economic Development Studies (CEDE) at Universidad de los Andes. It has information on royalties, municipal budgets, institutional presence of the central government, armed groups presence, and other characteristics of Colombian municipalities (Acevedo and Bornacelly, 2014). Finally, for Peru, we constructed a panel with analogous characteristics for each municipality.³⁰

We present summary statistics for municipalities in Colombia and Peru in Table 1 – Panels B and C, respectively. Mining municipalities in each country have a similar area, but there are more inhabitants in Colombian municipalities, and more area is covered by mining permits in Peru. We exclude from the analysis municipalities without mining potential in the subsoil, as there can be no mining in those municipalities.

4.1. Constructing the illegal mining panel

We have information from the 2010 Colombian Mining Census, published by the Ministry of Mines, on the locations of legal and illegal mines in Colombia. Unfortunately, we do not have the same information for Peru to train the model with Peruvian data. The census covered half of the Colombian municipalities (Table 1 – Panel D). We do not find any evidence that the municipalities sampled by the census were selected based on observable characteristics (Table A.4 in the Appendix provides more details). Municipalities included in the census are similar to those not included in terms of change in royalties due to the reform, production of different minerals, institutional presence of the central government, and presence of armed groups.

Most of the mines in the census are open pit. Hence, they can be observed with satellites and detected by our machine learning algorithm. We calibrate an algorithm with this information and use it to predict mining activity in other regions across the years. There are six main steps in constructing the panel of illegal mining by municipality: (i) Prepare the satellite imagery so it can be used in the prediction model. (ii) Calibrate a machine learning algorithm with the 2010 census data to predict whether an individual pixel is mined. (iii) Predict mining activity in all pixels for the years 2004 to 2014 with the model built in the previous step. (iv) Assess the legality of each mined pixel using the map of legal permits. (v) Identify the mineral mined in each location. To do this, we use the potential subsoil resources mapped by the National Mining Agency (Agencia Nacional de Minería, 2013). (vi) Collapse the results at the municipality-mineral level for the regression analysis.

To train the model in step (ii), we have the following information for each 30×30 m pixel (square) of the censused municipalities: a label denoting whether the pixel has mining activity, six satellite surface-reflectance measures for different bands (from NASA's LANDSAT 7 satellite), deforestation year (Hansen et al., 2013), and ecosystem type (Etter, 2006).³¹ We split the sample, allocating 75% of the observations for training (learning) and 25% for testing.

The goal of our machine learning algorithm is to detect the footprint of an open-pit mine (e.g., the white part in Fig. 2). One could impose a rule for declaring a pixel as mined or allow the machine to “learn” the optimal rule, based on the characteristics of known mines. For example, we could impose the following rule: Every pixel without forest, not located in a desert, and a color close to white is a mine. Instead, we let the computer try different nested binary decision rules: trees, as they are known in the machine learning literature.³² We aim to find a model with a high true positive rate (i.e., it labels true mined pixels as mined), but with a low false-positive rate (i.e., it does not label unmined pixels as mined). We expect the relationship between the existence of a mine and the satellite measurements to be highly nonlinear and complex; therefore, we use random forests, which are suitable for this type of problem (James et al., 2014). A random forest, as its name indicates, is a collection of many binary decision trees. But in each node, the candidate subset of explanatory variables to be used in the binary partition is random. We tried a logit model more familiar to economists, but it generated twenty times more false positives.

The random forest model attaches to each pixel in each year a probability that it is mined. We then need to determine the optimal cutoff at which we declare a pixel to be mined.³³ For each cutoff, we plot in Fig. A.2 the associated true positive rate (TPR) and false-positive rate (FPR) in the testing sample. Ideally, we want to have 100% TPR and 0% FPR (upper left corner). When we lower the cutoff, we detect more mined pixels (improving the TPR), but at the cost of miss-classifying more unmined pixels (higher FPR). In the machine learning literature, it is standard to choose the cutoff ρ such that $TPR(\rho) - FPR(\rho)$ is maximized (marked with a blue star in Fig. A.2). There are two important aspects of our analysis and data that make this standard cutoff less appropriate. First, we are using the predictions as the dependent variable. Second, our sample includes many unmined pixels. We discuss both issues in Appendix D.1 in Appendix. In a nutshell, the formula we use to choose the optimal cutoff (see Eq. (10) in Appendix D.1) minimizes the sum of squared errors in the generic treatment effect regressions we estimate. The formula weighs more having a lower FPR, given that most pixels in the country are not mined. Although we use this optimal cutoff in our main regressions, we also show that our estimates are robust to using the standard cutoff (that has a higher true positive rate).

The area under the curve of our prediction model is 87%, which is similar to that obtained with a convolutional neural network for infrastructure detection (Oshri et al., 2018). In Table 2 we present the confusion matrix for the optimally chosen cutoff. This

²⁹ <http://www.simco.gov.co/>

³⁰ The panel uses data from the National Statistics Bureau-Peru and the Peruvian Ministry of Mines. Specifically, we obtain population from <http://proyectos.inei.gob.pe/web/poblacion/> and mineral production by municipality from <http://mineria.minem.gob.pe/detalle-estadistica/?id=12501>.

³¹ LANDSAT 7 captures different wavelengths in different bands. Specifically, we use Band 1 (blue), Band 2 (Green), Band 3 (Red), Band 4 (Near-infrared), Band 5 (Shortwave infrared 1), and Band 7 (Shortwave infrared 2). These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. See <http://lpdaac.usgs.gov> for more details.

³² The name “tree” comes from the graphical representation of the nested binary decision rules.

³³ The downside of using the raw probabilities is that the measure of the fraction of area mined will be affected by the probability the model assigns to unmined pixels. As a robustness check, we present results using the raw probability. The results are qualitatively similar.



Fig. 2. Image of the footprint of a mine. *Notes:* Example of a mine we aim to detect with the machine learning algorithm. The white portion of the image is the mine footprint, in contrast to the river (brown) and vegetation (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Source: Digital Globe-Google Maps.

Table 2
Confusion matrix of the mining detection model.

	Non-Mined	Mined
Predicted Non-Mined	131,747	2,972
Predicted Mined	382	1,428

Notes: The confusion matrix presents the accuracy of the prediction model in classifying mined pixels using the optimal threshold. The columns show the actual mined status of the pixels according to the training data, while the rows show what the model predicts. The precision rate is 79% ; that is, of the pixels predicted to be mined (Predicted Mined row), 79% are actually mined.

matrix presents the number of correctly/incorrectly classified mined/non-mined pixels. The rate of precision is 79% . That is, of the pixels we predict as mined, almost four-fifths are truly mined according to the testing data. Our model correctly classifies 32.45% of true mine pixels (TPR), and wrongly classifies as mines 0.29% of pixels without a mine.³⁴ Classification errors add noise to our outcome variable, which decreases the precision of our estimates. However, since we do not see any evidence of differential classification error across treatment and control areas, the error is unlikely to induce bias in our estimates. Appendix D provides further details on the steps to construct the illegal mining panel.

4.2. From pixel predictions to municipality panel

After predicting whether a given pixel is mined each year, we overlay the map of legal permits of that year to declare the pixel as legally or illegally mined.³⁵ We obtain locations and exact shapes of Colombian legal mines from Tierra Minada.³⁶ The data for Peru was obtained from the Geological Mining and Metallurgical Institute of Peru.³⁷ To determine the mineral being mined in each Colombian pixel classified as mined, we use the mining potential map (i.e., what mineral can be expected in each location) produced by the National Mining Agency (Agencia Nacional de Minería, 2013).³⁸

³⁴ The TPR is similar (26%) when testing our model on the illegal gold mines manually identified by UNODC (2016).

³⁵ We smooth our predictions to prevent having pixels that switch back and forth from mined to unmined due to prediction error. We do this by calculating the monotonic sequence of 0's (unmined) and 1's (mined) closest to the vector of each pixel prediction through time.

³⁶ Tierra Minada is a nonprofit organization that digitized official records contained in the Catastro Minero Colombiano (Colombian Mining Cadastre). The full dataset can be downloaded from <https://sites.google.com/site/tierraminada/>.

³⁷ Accessed through Global Forest Watch on May 22, 2016. www.globalforestwatch.org.

³⁸ An attempt to predict the mineral mined in each area with a machine learning model was not sufficiently precise. See Fig. A.3 in Appendix.

Table 3
Summary statistics, illegal mining panel.
Source: Authors.

Panel A: Mineral mining Colombia			
% of mined area mined illegally	Coal	Gold	Difference
Before the reform	85.07 (26.75)	88.67 (24.75)	3.6*** (.66)
After the reform	79.88 (27.85)	83.95 (26.21)	4.07*** (.8)
Difference	-5.2*** (.82)	-4.73*** (.64)	.47 (1.03)
Panel B: All mining Colombia-Peru			
% of mined area mined illegally	Peru	Colombia	Difference
Before the reform	88.66 (20.56)	88.5 (23.4)	-.16 (.36)
After the reform	78.02 (26.94)	82.33 (25.55)	4.31*** (.55)
Difference	-10.64*** (.36)	-6.17*** (.53)	4.47*** (.63)

Notes: The Table presents summary statistics for the illegal mining panel we constructed using the machine learning model. Panel A presents the mean percentage of the municipality mined area that is illegally mined, by mineral for Colombia. In the columns, the results are presented by gold/coal, and in the rows, they are presented for the years before the reform (2004–2010) and after the reform (2011–2014). UNODC (2016) estimates for Colombia and our estimates of “% of the mined area mined illegally” are higher than the Census estimates presented in Table 1. Likely some illegal mines refused to answer the census or under-reported their mining area. Panel B presents the mean percentage of the whole municipality mining area mined illegally for Colombia and Peru. In the columns, the results are presented by country, and in the rows, they are presented for the years before the reform and after the reform.

Finally, we add the predictions at the municipality-mineral level. We calculate the percentage of the mined area of each mineral that is illegally mined. We create two measures: “cumulative” and “new”. Cumulative refers to all the mined pixels the model detects in the image of a year, regardless of when mining activity started. New refers to the pixels classified as mined, which had not been identified in the previous year.

Table 3 presents the summary statistics of our predictions. These statistics are also a preview of our two identification strategies: by rows, we separate before and after the reform; by columns, we have coal/gold (Panel A) or Peru/Colombia (Panel B). According to our estimates, 84% of the gold-mining area in Colombia is exploited without a permit after the reform. Although this number seems high, it is close to the 81% estimated for gold mining in 2014 using manual inspection by the UNODC (2016).

The final panel has data from 2004 to 2014. While there are satellite images covering more recent years, we do not have information on mining titles given in Peru after 2014. Thus, we are unable to assess the legality of mines after this period. In addition, on December 20 of 2014, the FARC guerrilla in Colombia announced a unilateral ceasefire, which marked the beginning of a continued de-escalation of the internal conflict in Colombia until the final peace agreement in 2016. The de-escalation of the conflict changed the illegal mining dynamic in the country (Rettberg and Ortiz-Riomalo, 2016; Baptiste et al., 2017; Massé and Billon, 2018). Therefore, we restrict the main analysis to the 2004–2014 period. However, we present in Figs. A.8 and A.9 results including two additional years, using 2014 titles to assess the legality of mines in 2015 and 2016.

5. The effect of the reform on illegal mining

5.1. Identification strategies

We use two different identification strategies to test the predictions of our model. A difference-in-differences strategy comparing minerals across time within the same municipality to test that illegal mining increased more for minerals with a higher tax rate. Furthermore, to test that mines are more likely to operate illegally after the reform in Colombia, we use a difference-in-differences strategy that compares illegal mining in Colombia and Peru, before and after the reform. We then use cross-sectional variation in municipality characteristics to identify the heterogeneous effect of the reform on municipalities with weak central government presence (Prediction 3) and illegal armed groups presence (Prediction 4).

To test Prediction 1 — that the reform increased illegal mining — we use a difference-in-differences framework comparing Colombia and Peru, before and after the reform. We cannot use within Colombia variation because all the Colombian municipalities were affected by the change in incentives. We use Peru as the control for several reasons: (i) It is a neighboring country that is also affected by illegal mining; (ii) Peru has levels of mineral production similar to Colombia according to a hierarchical clustering analysis (see Fig. A.4 in the Appendix); and (iii) Peru and Colombia simultaneously adopted an Andean Community law allowing

the destruction of illegal mining machinery on-site.³⁹ The estimating Equation is:

$$\widehat{y}_{mt} = \beta_A After_t \times Col_m + \gamma_m + \gamma_t + \varepsilon_{mt}, \quad (4)$$

where \widehat{y}_{mt} is the estimated percentage of the area mined that is mined illegally in municipality m in year t . We focus on the share of the area mined illegally for two main reasons. First, because it is the variable for which the theoretical model has predictions. Second, because the share avoids confounding possible overall changes in total mining activity. $After_t$ is a dummy variable equal to one after the royalty reform (i.e., for $t \geq 2011$). Col_m is a dummy variable indicating whether the municipality is Colombian. γ_m and γ_t are municipality and year fixed effects, respectively. The municipality fixed effects capture time-invariant characteristics of the municipalities, like mineral resources and distance to the capital. While the time fixed effects capture common shocks each year to both countries, like the financial crisis of 2008 or fluctuations in the international price of minerals. ε_{mt} is an error term that we cluster at the municipality level.

The identification for β_A in Eq. (4) comes from changes in illegal mining before and after the reform, in Colombia compared to Peru. If other events that affect mining took place around the same time as the reform, β_A would confound the effect of those events with the effect of the reform. Of particular concern here are two such concurrent events. First, in Colombia, the central government's system for processing mining-permit requests was down around the time of the reform. Although one might expect that the firms that wanted to get a permit would wait or come into compliance once requests were being accepted again, we cannot fully separate these two effects. To address this concern, we define illegal mining areas as mining areas outside the legal permits at the end of the study period, 18 months after the system was working again. That is, if a miner could not register a permit while the system was down (in 2012 and 2013) but acquired a permit within the next 18 months (by the end of 2014), it will not count as illegal mining in our data.⁴⁰

The second event simultaneous with the reform was a change in the law allowing the destruction of illegal mining machinery on-site, rather than confiscation and a court procedure. Both countries approved this law, and hence the difference-in-differences estimator will not be affected. However, a potential confounding factor could stem from a differential degree of enforcement of this law. While we do not have data on the destruction of machinery in the two countries to study this question directly, the estimates from the within-municipality analysis used to test for heterogeneity across minerals (Prediction 2) are not affected by this concern.

Another issue that might bias β_A is that the machine learning model used to detect mining was trained using only data from Colombia. Thus, it might be less accurate to identify mines in Peru. If the dependent variable used is the area of illegal mining and the area of illegal mining in Peru is systematically underreported by the algorithm, then the estimate of the reform would be biased.⁴¹ However, since we use the percentage of mined area mined illegally as the dependent variable, this bias does not arise as long as detection rates across legal and illegal mines are the same (for each country, in each year). The difference-in-differences estimate is not affected since both the numerator and the denominator are affected by the differential measurement in both countries, which leaves the proportion of illegal mining unaffected.

Fig. 3 shows that Colombia and Peru had similar trends and levels of illegal mining before the reform. We also test the robustness of our results following Rambachan and Roth (2020) "honest difference-in-differences" approach. Overall, our results are robust to plausible violations of the parallel-trend assumption (see Fig. A.5).

To test Prediction 2 — the increase in illegal mining is larger for minerals with a larger royalty rate — we use an identification strategy that relies on variation in the increase of illegal mining for different minerals within the same municipality in Colombia. For example, given that gold has a higher royalty rate than coal, we expect a larger increase in the share of illegal mining in gold-mining areas compared to coal-mining areas in the same municipality. We use municipality-time fixed effects that capture time-varying municipality events, such as a new mayor taking office and the peace process; as well as nationwide changes, like the permit-request system being down. The mineral-municipality fixed effects capture constant differences in the production function of extracting different minerals in each municipality. Since we identify the difference in illegal mining by mineral in each municipality, the sample does not include municipalities with only one type of mineral. Our preferred specification will only use gold and coal, as they are the most mined minerals according to the mining census (Table A.5). We include in the Appendix other minerals as robustness. Specifically, we estimate the following Equation:

$$\widehat{y}_{imt} = \beta After_t \times \alpha_i + \gamma_{mt} + \gamma_{im} + \varepsilon_{imt}, \quad (5)$$

where \widehat{y}_{imt} is the estimated percentage of area mined illegally for mineral i , in municipality m at time t . α_i is the royalty rate of mineral i . $After_t$ indicates after the royalty reform. Finally, γ_{mt} and γ_{im} are municipality-time and mineral-municipality fixed effects. Fig. 4(b) suggests that the parallel trends assumption is also satisfied for this difference-in-difference strategy. We also test the robustness of our results following Rambachan and Roth (2020) "honest difference-in-differences" approach. While these results are more sensitive to plausible violations of the parallel-trend assumption (see Fig. A.6), this is likely due to small sample sizes and sample variability when using tight fixed effects.

³⁹ Brazil was another candidate to serve as a control, but it is not a member of the Andean Community, and thus is not affected by the law allowing the destruction of illegal mining machinery on-site. Venezuela and Ecuador have similar levels of mineral production but do not have maps/shapefiles of legal mining permits available.

⁴⁰ Mining permits were not accepted between February 2, 2011, and July 2, 2013. This resulted from Law 1382 of 2010, which mandated an upgrade of the platform used to grant and archive permits. This interruption is unlikely to have had an immediate effect on illegal mining, due to the time lag between a permit request and its granting or denial (which averaged over one year before the interruption). Appendix B.2 provides more details.

⁴¹ Although we do not have 2010 Peruvian mines data to train the model, we found some data on the location of 213 mined pixels in the Peruvian Amazon in August 2020 (<http://rami.servirglobal.net/>). When we compare the 2020 Peruvian mines with our predictions for 2014, the true positive rate is 20% smaller than in Colombia.

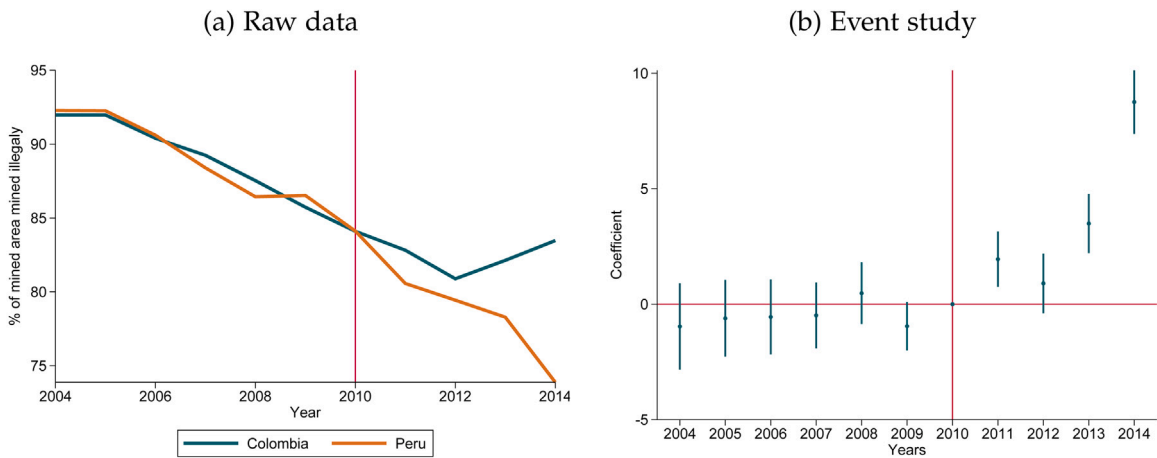


Fig. 3. Parallel trends between Colombia and Peru. *Notes:* Both panels use municipality-year data from 2004 to 2014 for Colombia and Peru. Fig. 3(a) shows the evolution of the average percentage of mined area mined illegally in each country from 2004 to 2014. The estimates in Fig. 3(b) are from an event study regression for the percentage of mined area mined illegally. The x-axis plots time in years and the y-axis the coefficient of the indicator of Colombia interacted with the respective year. Point estimates and 95% confidence intervals are plotted. 2010 is the base year. The drop for the 2012 coefficient is due to our conservative assumption of defining illegal mining areas in 2012 and 2013 (while the permit system was down) as mining areas outside the legal permits at the end of 2014, 18 months after the system was working again. See Fig. A.10 for a version of this graph without this assumption, i.e. using concurrent permits.

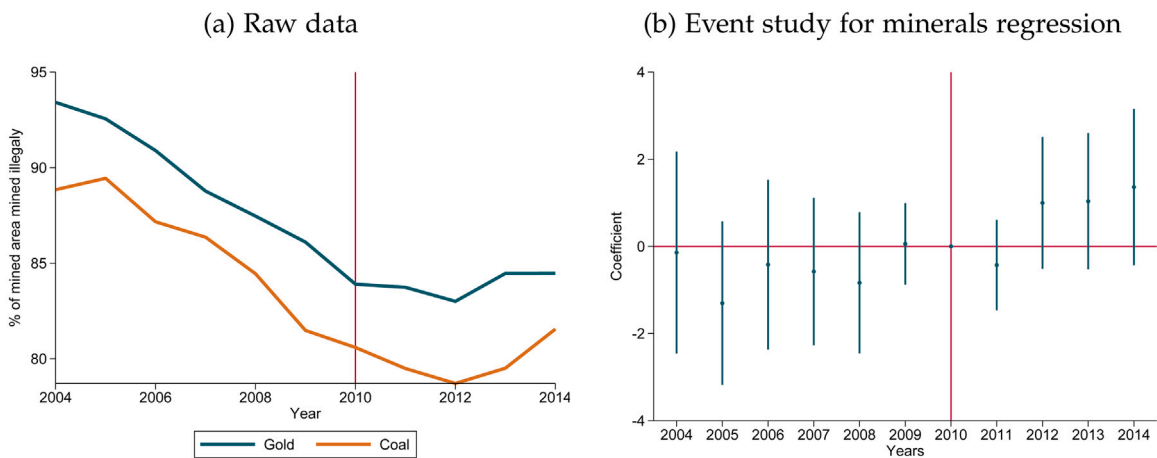


Fig. 4. Parallel trends between Gold and Coal. *Notes:* Both panels use municipality-mineral-year data from 2004 to 2014 for Colombian municipalities. Fig. 4 shows the evolution of the average percentage of mined area mined illegally across municipalities for gold and coal mining in each year. The estimates in Fig. 4(b) are from an event study regression for the percentage of mined area mined illegally by mineral. The x-axis plots time in years and the y-axis the coefficient of the royalty rate interacted with the respective year. 2010 is the base year. Point estimates and 95% confidence intervals are plotted.

To study heterogeneity of the reform on municipality characteristics (i.e., Predictions 3 and 4), we estimate equation (4) and interact the dummy of after the reform in Colombia with the given characteristic Z_m .⁴²

$$\widehat{y}_{mt} = \beta_{AZ} After_t \times Col_m + \beta_Z After_t \times Col_m \times Z_m + \gamma_m + \gamma_t + \varepsilon_{mt}, \tag{6}$$

5.2. Prediction 1: Effect on the share of illegal mining

We first study how the difference between Colombia and Peru evolves before and after the reform (Fig. 3), using an event study (or dynamic difference-in-differences). As mentioned above, Colombia and Peru had similar levels (and trends) of illegal mining before the reform, but illegal mining is higher in Colombia (relative to Peru) after the reform. The percentage of the mined area mined illegally was falling similarly in both countries before the reform. Nevertheless, after the reform it continued falling in Peru, but in Colombia it increased slightly.

⁴² We do not have data for Peru on Z_m , so we cannot include $After_t \times Z_m$. As a result, the interaction coefficients need to be interpreted as the effect relative to the average municipality in Peru (as opposed to relative to a municipality with the same characteristics).

Table 4
Effect of the reform on illegal mining.

Dependent variable:	% mined area mined illegally		Area mining permits (km ²)
	Cumulative (1)	New (2)	Cumulative (3)
After x Colombia	4.21*** (0.60)	5.12*** (0.67)	-11.8*** (2.80)
N. of obs.	26,323	10,679	26,323
Colombian municipalities	904	741	904
Peruvian municipalities	1797	714	1797
Mean of dep. var.	85.1	88.1	46.7
R ²	0.73	0.73	0.87
Within R ²	0.0058	0.012	0.0038

Notes: Estimated coefficients using Eq. (4) with different dependent variables. The estimations use municipality-year data from 2004 to 2014 for Colombia and Peru. In the first column, the dependent variable is the percentage of the cumulative area mined that is mined illegally in the municipality. Hence, Column 1 only includes municipalities with detected mining. In Column 2, we calculate this percentage only in the new area mined that year. The number of observations is smaller since the estimation requires a cloud-free image in consecutive years; also, we cannot use observations from 2004 because we do not have satellite images from the previous year. In the last column, the dependent variable is the total area of mining permits, measured in square kilometers. All regressions include municipality and time fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 presents the results of estimating equation (4). As predicted, the area mined illegally as a share of the total mined area increased by 4.21 percentage points in Colombia relative to Peru, after the reform (Column 1). The effect of the reform is larger when we focus on the fraction of newly mined area that is mined illegally (Column 2 in Table 4), because this measure excludes the existing mines (which may be less sensitive to the reform). Another way of confirming our results is to estimate an analogous regression using the area covered by mining permits as the dependent variable (Column 3 in Table 4). This measure does not depend on our mining-area predictions and is calculated from the government's data. We find a reduction, relative to Peru, in the area covered by mining permits in Colombia after the reform. As mentioned above, the Colombian government did not accept mining permit requests between February 2, 2011, and July 2, 2013, which could partially explain the reduction in the area covered by mining permits after the reform. Fig. A.7 shows the evolution of mining permits in both countries before and after the reform. The relative reduction in mining permits in Colombia persists beyond 2013, suggesting the reduction is driven by an increase in illegal mining as opposed to the inability to obtain permits. In addition, if we remove the years when the government did not take mining permit applications, the results are very similar and suggest we may be underestimating the effect of the reform in the full sample (see Panel E in Table A.6). Since the royalty reform was approved in mid-2011, we do two alternative specifications: One in which we define the treatment as starting in 2012 (as opposed to 2011), and one in which we remove the data from 2011 from the sample. The results are robust to both specifications and the treatment effects are qualitatively similar (see Table A.6). The results are also robust to extending the data (with some assumptions, see Section 4.2) to include the years 2015 and 2016 (see Fig. A.8).

Besides addressing the concern that permit requests were not accepted in Colombia at the time of the reform, we address two other issues. First, the change in the royalty transfers caused by the reform may alter the capacity of enforcement from the local governments. However, as mentioned above, the local police and the national army are funded by the central government. When we control for changes in the amount of royalty transfers, the coefficient is qualitatively similar (see Columns 2–4 in Table 5).

Second, because the machine learning model used to detect mining was trained using only data from Colombia, it might be less accurate identifying mines in Peru. While we do not have data to validate performance across countries or through time, we can test whether within Colombia the effect of the reform is higher in municipalities used to train the model. We find that the effect of the reform is smaller in municipalities used to train the machine learning model (see Column 5 in Table 5). Thus, it is unlikely that we observe more illegal mining after the reform in Colombia because of bias induced by the data used to train the machine learning algorithm.

Robustness

In Tables A.7–A.8 we present additional robustness checks of these results. We first assess whether our results are robust to controlling for other covariates in the regression. As the set of possible controls is large, we use a Double Lasso procedure to select relevant controls from a statistical point of view and are not chosen ad hoc by us. The Lasso procedure is like an ordinary least squares regression where the sum of squared residuals is minimized, but there is also a penalty for the number of controls used (James et al., 2014). In the set of possible candidates, we include a price index incorporating the price of different minerals, population, and homicides by armed groups, and these variables squared, lagged, interacted among them, interacted with a linear trend, and interacted with a quadratic trend. We use the Stata program provided by Belloni et al. (2014) to implement their Double Lasso procedure (Table A.7 in the Appendix provides the results). The procedure selects, among others, the lagged price, which is intuitive given the time it takes to start a mining operation. The coefficient of “After x Colombia” is 10% smaller when including the optimal controls, but still significant at the 1% level.

In the theoretical model, we assumed the miner could become compliant if he paid the permit fees and royalties. However, not all illegal mines can acquire a mining permit. For example, mining inside a national park is not allowed, and hence permits within

Table 5
Heterogeneous effects of the reform.

Dependent variable:	% mined area mined illegally						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After x Colombia	4.13*** (0.60)	3.89*** (0.66)	4.38*** (0.57)	4.25*** (1.04)	6.06*** (0.76)	3.46*** (0.68)	4.55*** (0.70)
After X Col X Loser		1.37 (0.98)					
After X Col X % Budget Loss			0.071*** (0.025)				
After X Col X % Budget Loss if Loss				0.077** (0.034)			
After X Col X % Budget Win if Won				-0.054 (0.12)			
After X Col X Censed					-3.36*** (0.93)		
After X Col X Weak Institutions						2.74*** (0.96)	
After X Col X Armed Groups							-1.05 (0.97)
N. of obs.	25,982	25,982	25,982	25,982	25,982	25,982	25,982
Colombian municipalities	867	867	867	867	867	867	867
Peruvian municipalities	1797	1797	1797	1797	1797	1797	1797
Mean of dep. var.	85.1	85.1	85.1	85.1	85.1	85.1	85.1
R ²	0.73	0.73	0.73	0.73	0.73	0.73	0.73
Within R ²	0.0055	0.0057	0.0059	0.0059	0.0069	0.0062	0.0057

Notes: The entries in the table are the estimated coefficients of Eq. (4) with the added heterogeneity in each case. The estimations use municipality-year data from 2004 to 2014 for Colombia and Peru. All regressions include municipality and year fixed effects. Column 1, repeats the specification of Table 4 Column 1, but restricted so that all columns have the same number of observations. % budget loss refers to the negative change in royalties received by the municipality 2013–2014 compared to 2010–2011, as percentage of the municipality budget 2010–2011. If the municipality has more royalties the loss is negative. “Loser” is a dummy for whether the municipality receives less royalties transfers after the reform. Column 4 separates those with negative change (“loss”) and positive change (“won”). Censed is a dummy indicating if the municipality was part of the 2010 Mining Census. Weak institutions is a dummy that takes the value of 1 if the municipality has a below the median number of institutions (e.g., tax collection or notary office) per capita (Acevedo and Bornacelly, 2014). We measure illegal armed group presence with a dummy indicating whether the municipality had any reported homicides committed by an armed group (Acevedo and Bornacelly, 2014). Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

park boundaries should not be issued. Thus, we expect (and find) that the increase in illegal mining is concentrated outside national parks (Table A.8, Column 2). The results using the raw probabilities that a pixel is mined, instead of a dummy, are qualitatively similar (Table A.8, Column 3).

We also study whether there are spillovers across the border (i.e., whether the royalty reform in Colombia affected mines near the border in Peru). To do this, we estimate a regression model including an interaction with a dummy for municipalities more than 1,000 km away from the border (Table A.8, Column 4). Since we cannot reject that the interaction coefficient is zero, we take this as evidence that spillovers are not driving the result. We also investigate whether the results are robust to using a different cutoff for the machine learning predictions (Table A.8, Column 5). In particular, we use the point closest to the ideal of correctly predicting all mines (100% TPR) and no false positives (0% FPR). For our model, it is a cutoff associated with an 80% TPR and a 20% FPR. The magnitude of the estimated coefficient is almost double the coefficient with the optimal threshold. This increase is because the new cutoff has almost double the difference between TPR and FPR, compared to our conservative optimal threshold.⁴³

Finally, weighting the observations by the fraction of the municipality area that is analyzed does not change the nature of the results (Table A.8, Column 7). In short, the qualitative nature of the result is robust to different specifications (Table A.8).

So far, we have looked at the extensive margin of evasion. Still, evasion may also be present on the intensive margin through underreporting to the central government of quantity produced by legal mines. We estimate equation (4) using reported production per area as the dependent variable (see Table 6). There is no positive (increase in reported quantity) significant effect on any mineral (p -values is 74% for gold). At the same time, there is a negative (reduction in reported quantity) and significant treatment effect of silver (p -value 0.05) and a negative but insignificant effect for coal (p -value 0.11). Thus, there is some weak evidence of an increase in underreporting. The lack of a larger effect on the intensive margin could be explained by the difficulty monitoring the quantity extracted compared to the area mined. The local authority and the miner likely bargain over the mine’s area, which is what the former can observe. Although the magnitude of the coefficient for silver is large relative to the mean, we assume, conservatively, that there is no increase in underreporting when monetizing the increase in evasion with the reform.

⁴³ Sixty percent compared to 32%. The difference between the TPR and the FPR is directly proportional to the estimate of the area illegally mined. Adjusting the raw measure of the fraction of the municipality area that is mined, using the formula in Eq. (9), yields similar results (Table A.8, Column 6). Appendix D.1 provides more details.

Table 6
Effect of the reform on reported quantity.

Dependent variable: Reported production by area	Gold		
	Coal (1)	(2)	Silver (3)
After x Colombia	-3.33 (2.02)	6.33 (19.0)	-9.26** (4.62)
N. of obs.	791	2,066	1,976
Colombian municipalities	101	205	177
Peruvian municipalities	13	104	121
Mean of dep. var.	3.92	20.7	4.42
R^2	0.33	0.29	0.27
Within R^2	0.00033	0.000072	0.0033

Notes: Coefficients estimates of Eq. (4), using as the dependent variable the reported production of each mineral by area. The estimations use municipality-year data from 2004 to 2014. All regressions include municipality and time fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3. Prediction 2: Effect of the reform by mineral and royalty rate

The results of estimating Eq. (5), the differential effects of the reform by mineral, are presented in Table 7. Column 1 confirms Prediction 2: For a mineral with a royalty tax rate one percentage point higher, there was an extra 1.18 percentage point increase in the area illegally mined. Since the royalty reform was approved in mid-2011, we do two alternative specifications: One in which we define the treatment as starting in 2012 (as opposed to 2011), and one in which we remove the data from 2011 from the sample. The results are robust to both specifications and the treatment effects are qualitatively similar (see Table A.9).⁴⁴ The results are also robust to extending the data (with some assumptions, see Section 4.2) to include 2015 and 2016 (see Fig. A.9).

The results from Tables 4 and 7 are difficult to compare. Besides assumptions about external validity and imposing a linear-dose relationship, one would need to assume that the results from Table 7 — which estimates the effect of different royalty rates after the reform takes places — apply to Peru, even in the absence of a royalty reform there. However, both identification strategies support each other in the predictions of the model.

We also study the evolution of illegal mining by mineral (Fig. 4(b)) using an event study (or dynamic difference-in-differences). As mentioned above, the trend of illegal mining in gold-bearing areas is parallel to that in coal-bearing areas before the reform, but it diverges immediately after the reform. We try other illegal mining measures analogous to the ones we did for the Colombia-Peru differences-in-differences. Column 2 of Table 7 presents the results with new area mined and Column 3 with the area in square kilometers that is covered by a mining permit. The results for the new area mined are in line with the main results, suggesting an increase after the reform is passed. However, we do not observe any difference in the area covered by mining permits.

Robustness

Table 8 presents robustness exercises similar to those in Section 5.2. The coefficient is similar and statistically significant when excluding national parks (Column 2), using the continuous probability that a pixel is mined (Column 3) or including weights (Column 6). However, when changing the cutoff for declaring a pixel as mined (Column 4) or adjusting the predictions to account for false positives (Column 5), the results are not statistically different from zero. These can be explained by the fact that the area with mining potential is small, and hence the false positives have more incidence on the calculations. Finally, we present robustness to including other minerals on the regression in Table A.10. Recall that our preferred estimation only uses gold and coal mines, which are the two minerals mined the most according to the mining census (Table A.5). The results are robust to including platinum and magnesium.⁴⁵

5.4. Prediction 3: Effect of the reform in municipalities with lower national oversight

The theoretical framework predicts a larger effect of the reform in municipalities with a low probability of illegal mines being detected independently by the central government. Municipalities with a low probability of detection are those with weak institutional presence, measured as the number of central government institutions (e.g., tax collection or notary office) per capita (Acevedo and Bornacelly, 2014). As predicted, the increase in illegal mining is larger in areas with weak institutional presence, almost 80% larger $((3.46 + 2.74)/3.46)$ (see Table 5, Column 6).

⁴⁴ A possible concern is that the Andean Community law allowing the destruction of illegal mining machinery on-site after 2012 explains the change in illegal mining across minerals. However, in most cases, the police illegal-mining actions are against illegal gold mining. For example, we obtained information from all the police operations against illegal mining conducted between January 2014 and August 2014. Of these, eight times more actions correspond to gold mining sites compared to coal. The reform made these actions more effective and thus, if anything, affected gold mining more.

⁴⁵ The map of mining potential (Fig. D.7) has no information for the other most-mined minerals of Table A.5 like clay or sand.

Table 7
Effect of the reform on illegal mining by mineral.

Dependent variable:	% mined area mined illegally		Area mining permits (km ²)
	Cumulative (1)	New (2)	Cumulative (3)
After $\times \alpha$	1.18** (0.58)	1.29* (0.76)	3.41 (2.19)
N. of obs.	6,774	4,918	6,774
Municipalities	390	378	390
Mean of Dep. Var.	83.85	86.47	10.63
R ²	0.96	0.90	0.90
Within R ²	0.0030	0.0016	0.0039

Notes: The entries in Table 7 are the coefficients estimates of Eq. (5), where the dependent variable is the percentage of the area mined that is mined illegally in the municipality region with potential to mine a given mineral. Area measured in square kilometers. The estimations use municipality-mineral-year data for Colombian municipalities from 2004 to 2014. All regressions include municipality-time and municipality-mineral fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8
Robustness of the illegal mining by mineral results.

Dependent variable:	% of mined area mined illegally					
	All (1)	Without Nat Parks (2)	Prob (3)	Cutoff (4)	Adjusted (5)	Weights (6)
After $\times \alpha$	1.18** (0.58)	1.05* (0.57)	1.36** (0.57)	0.095 (0.27)	-0.24 (0.48)	0.97* (0.58)
N. of obs.	6,774	6,760	6,866	6,774	2,014	6,772
Municipalities	390	389	395	390	156	390
Mean of Dep. Var.	83.9	83.8	83.9	91.1	80.7	83.8
R ²	0.96	0.96	0.96	0.98	0.99	0.96
Within R ²	0.0030	0.0024	0.0039	0.00015	0.00052	0.0020

Notes: The entries in Table 8 are the estimated coefficients of Eq. (5). The estimations use municipality-mineral-year data for Colombian municipalities from 2004 to 2014. The dependent variable is the percentage of the mined area that is mined illegally in the municipality for a mineral. Column (1) repeats the main specification. Column (2) excludes mined areas in national parks, which cannot be legalized even if the permit fees were paid. Column (3) uses the probability that a pixel is mined instead of the dummy of mined. Column (4) changes the cutoff of the mining predictions. It uses the cutoff where the difference between the true positive rate and false positive rate is maximized. Column (5) adjusts the measure of area mined according to Eq. (9). Column (6) weights each observation by the fraction of the municipality analyzed (i.e., cloud-free). All regressions include municipality-time and municipality-mineral fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5. Prediction 4: Effect of the reform in municipalities with illegal armed groups

We expect a smaller effect of the reform in municipalities with illegal armed groups because the local authority has little or no bargaining power. We measure presence with a dummy indicating whether the municipality had any reported homicides committed by an illegal armed group (Acevedo and Bornacelly, 2014). While this is an imperfect measure of the control of illegal armed groups, it is an objective and measurable proxy.⁴⁶ The effect is smaller, although not statistically significant (Table 5, Column 7). An alternative explanation would be that the National Police has targeted their efforts against illegal mining in areas where illegal armed groups get financial backing from this activity.⁴⁷

5.6. Dollars lost to evasion after the reform

After evaluating the four predictions, we estimate the dollars lost through evasion with the reform in three steps. First, we convert our coefficient of the effect of the reform into area mined illegally. Then we calculate the dollars lost in royalty taxes, and finally the dollars lost in permit fees. The coefficient of “After $\times \alpha$ ” is 1.18 (Column 1, Table 7). Multiplying this coefficient by the area mined and each mineral α_i , we obtain the increase in the area mined illegally due to the reform. Then we estimate the amount

⁴⁶ Attacks are only used while in military conquest/defense and do not necessarily reflect presence. For example, Puerto Boyacá, a town well known for being under paramilitary control for many years, did not have a single attack by the paramilitary while they controlled the area.

⁴⁷ Our request for longitudinal data on National Police operations to study these conjectures was denied.

Table 9
Data to estimate dollars lost with the reform.

Mineral (1)	Unit (2)	Price/unit (3)	Units/ha (4)	Alpha (5)	\$lost/ha (6)	Total ha (7)	Ha increase (8)	M \$ lost (9)
Gold	kg	44000	14.77	6%	38,992.80	76,894.92	5,444.10	212.28
Coal	ton	60	864.75	5%	2,594.25	39,077.82	2,305.54	5.98

Notes: Column (3) prices from Colombia's Central Bank and Mining and Energy Planning Unit. Data on Column (4) from the 2010 Colombian Mining Census. Column (5) is the royalty tax of each mineral. Column (6) is the product of columns (3)–(5). Data from Column (7) is estimated with our machine learning model. Column (8) is obtained by multiplying Columns (5) and (7) and the coefficient from [Table 7](#) Column (1). Column (9) is the product of Columns (6) and (8) in millions.

of mineral extracted per hectare from the 2010 Colombian Mining Census. Multiplying this by the royalty rate and price of each mineral, we obtain the revenue lost in royalty taxes. For gold, it is 212 million USD per year; for coal, it is 6 million USD per year. Thus, at least 218 million USD of royalty revenue is lost per year after the reform. Compared to the mining royalties distributed — 594 million USD in 2015 — this is equal to 37 cents per dollar redistributed. See [Table 9](#) for further details.

5.7. The optimal share of taxes for the local municipality

We motivated the paper with the problem of a fair distribution of resources and the transfer of tax revenue from rich to marginalized areas. The problem of the central government is to distribute one dollar of possible tax revenue between municipalities generating the revenue and other municipalities. A greater share allocated to the mining municipalities leads to smaller evasion and more resources to be distributed (as per our results). However, the central government might prefer to allocate resources to non-mining municipalities because they are poorer or because spending in the non-mining municipality might be more efficient. Using the increase in the share of the area mined illegally with the reform, we make a rough attempt to calculate the optimal share of taxes that should be distributed to the revenue-generating municipality.

Suppose spending is more efficient/preferred in mining municipalities. In that case, it is optimal to allocate all the mining taxes to the mining municipalities, since evasion is minimized and spending is more efficient/preferred there. However, [Gallego et al. \(2020\)](#) shows no differential effects per dollar spent in mining and non-mining municipalities in Colombia. Since spending is not more efficient in non-mining municipalities, the government must prefer to spend tax revenue there. We estimate that spending in non-mining municipalities would have to be three times as preferred for the reduction in tax revenue with the reform to be optimal.⁴⁸ Three times as preferred seems large, so there should be a larger share for mining municipalities. Indeed, in December 2019, a new law increased the share of royalties for mining municipalities in Colombia ([La Republica, 2019](#)).

6. Conclusions

In this paper, we study a reform in Colombia that reduced the share of tax revenue allocated to mining municipalities. The reform dramatically lowered the revenue local governments receive from legal mining in their territory and, consequently, their incentives to report illegal mining. Studying tax evasion and illegal activities is difficult as, almost by definition, these activities are hard to observe, and data is often scant and unreliable. We overcome this obstacle by using machine learning algorithms applied to satellite data to measure illegal mining over time.

As a share of total mined area, illegal mining increased in Colombia by 4.21 percentage points after the reform relative to the change observed for Peru. Across minerals, we find that for every percentage point of the royalty tax rate, there was 1.18 percentage point increase in illegal mining after the reform. Thus, for every dollar of royalties redistributed to the central government (as opposed to local governments), around 37 cents are lost through evasion. The increase in illegal mining illustrates the difficulties of redistributing resources. Given the trend towards decentralized spending ([OECD, 2018](#)), our results point to the importance of connecting tax revenue and spending. In short, there is a trade-off between efficiency and equity. While equity concerns may push for spending in certain locations (e.g., marginalized areas), this may result in lower revenue as local authorities become more lenient with evasion.

Our results also show that monitoring illegal activity using remote sensing and machine learning is promising. Indeed, many countries have started doing so. For example, India announced a policy along these lines and there are ongoing efforts in Colombia and Peru.⁴⁹ However, illegal miners could respond by resorting to underground mining or by disguising their mines, rendering monitoring more difficult. In that case, aligning the incentives of local bureaucrats is even more important.

Appendix A. Additional figures and tables

See [Figs. A.1–A.10](#) and [Tables A.1–A.10](#).

⁴⁸ See [Appendix C.3](#) for details of the calculation.

⁴⁹ http://articles.economicstimes.indiatimes.com/2016-04-12/news/72266895_1_minor-minerals-major-minerals-sand-mining

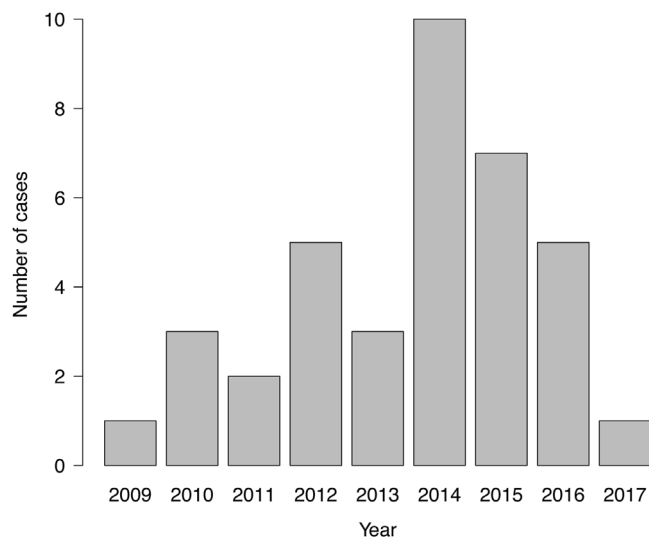


Fig. A.1. Disciplinary procedures related to illegal mining against mayors and governors by year. *Notes:* This figure presents the number of disciplinary procedures related to illegal mining opened by the Inspector General's Office (*Procuraduría General de la Nación*) against mayors and state governors between 2009 and 2017. The data was requested and given during 2017. Thus, the data does not include all the procedures opened in 2017.

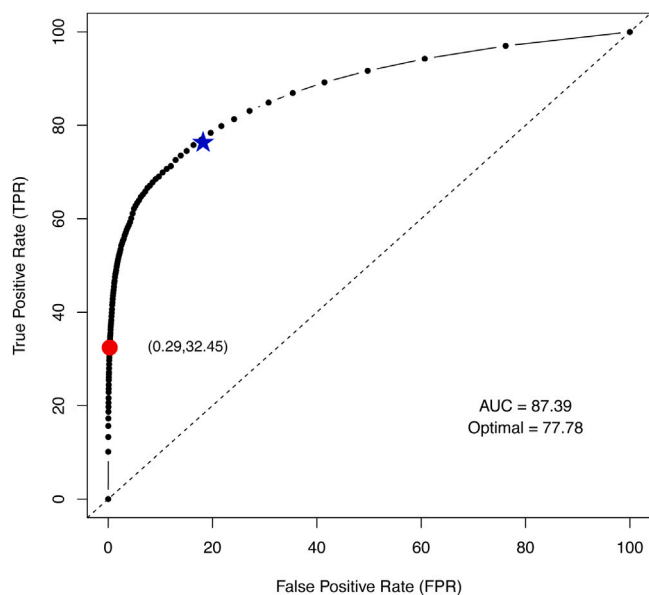


Fig. A.2. ROC curve for the mining prediction model. *Notes:* The receiver operating characteristic (ROC) curve plots the performance of a binary classification model when varying the cutoff threshold. The false positive rate (FPR)—the percentage of true no-mined pixels incorrectly classified as mined pixels—is on the x -axis. The true positive rate (TPR)—the percentage of correctly classified true mined pixel—is on the y -axis. As we decrease the cutoff to declare a mine, we accurately classify more true mined pixels as mined, but also increase the number of no-mined pixels incorrectly classified as mined.

Appendix B. Additional institutional background

B.1. Additional details of the reform in Colombia

The royalty reform in Colombia is outlined in Legislative Act 5 of 2011 (a constitutional reform) and Law 1530 of 2012. However, most of the motivation for the reform can be found in [Echeverry et al. \(2011\)](#). That paper was written by Colombia's Minister of Finance and Public Credit at the time (a leading advocate of the reform), the Director of Macroeconomic Policy at the same ministry, and the Director of Royalties at the National Planning Agency. The abstract reads:

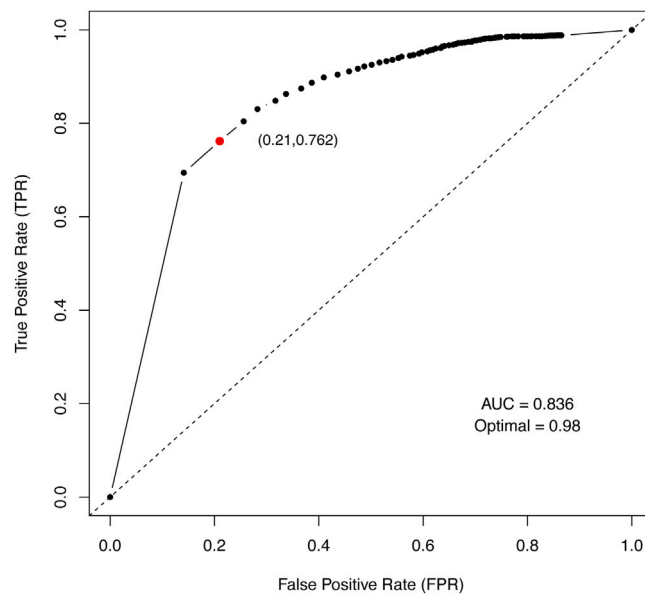


Fig. A.3. ROC curve for the mining prediction model trained to predict gold mining. *Notes:* See Fig. A.2 notes for explanation on the ROC curve. When we train the model using UNODC data only for gold mines the classification is not as good as our overall model. Classifying some gold-mined pixels as mines immediately misclassifies non-gold-mined pixels as mined. In other words, the FPR is high, and the formula of optimal cutoff obtains that it is best not to do any prediction unless less weight is assigned to the FPR.

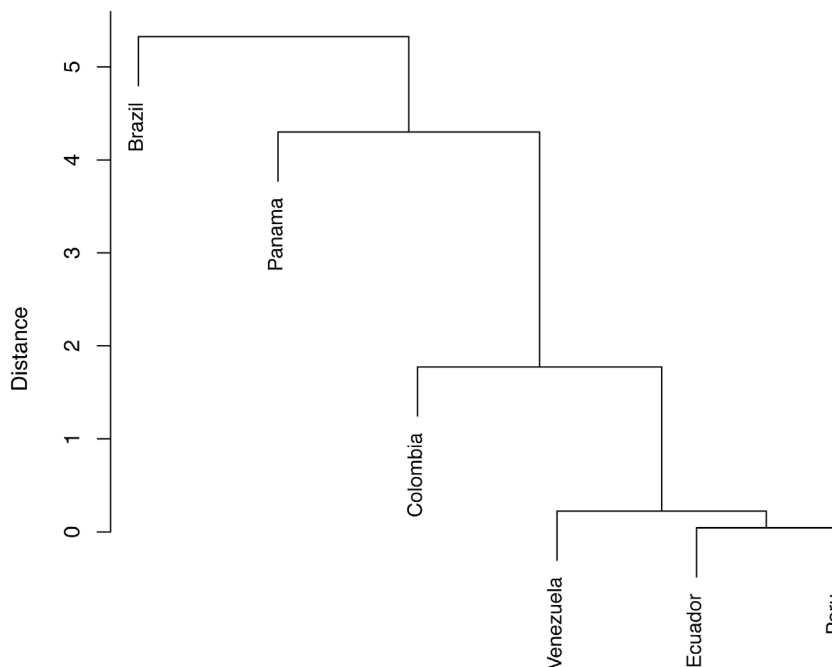


Fig. A.4. Hierarchical clustering of neighbor countries by mineral production. *Notes:* This dendrogram illustrates the distance between the normalized production vectors of each country from Table A.3. The vertical axis indicates how dissimilar the productions of two countries are. For example, Brazil's productions is relatively different from the Andean countries. Venezuela and Ecuador have similar levels of mineral production to Colombia, but they do not have maps of legal mining permits available.

“This policy memorandum accompanied the bill for Constitutional reform presented to Congress, aimed at creating the Royalties General System, which is based on four principles: (1) social and regional equality; (2) savings for the future; (3) regional competitiveness and (4) good governance.

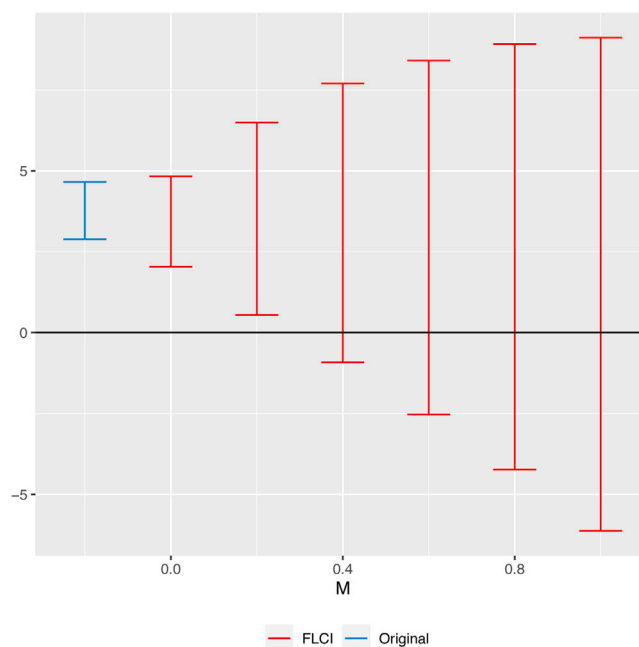


Fig. A.5. Robustness of the difference-in-difference estimates of Prediction 1 to violations of the parallel trends assumption. *Notes:* This figure examines how the robust confidence sets for the difference-in-difference estimates in Table 4 change as the parameter M (in $\Delta(M)$) varies following Rambachan and Roth (2020). M measures the change in slope of the differential trend. Typical difference-in-difference estimates assume $M = 0$ (i.e., that there is no differential trend in the absence of treatment). The mean change in slope in the pre-period in our data is 0.57, which roughly corresponds to the maximum value M can take where our results are still statistically significant at the 95%.

Table A.1
Royalties and municipality share by mineral.

Mineral	Royalty tax (α)	Municipality share (β)
Clay	1%	67%
Coal (0-3 tons)	5%	45%
Coal (>3 tons)	10%	32%
Construction materials	1%	67%
Copper	5%	40%
Emeralds	1.5%	Variable ^a
Gemstones	1.5%	Variable ^a
Gold	4%	87%
Gold (alluvial)	6%	87%
Gravel	1%	67%
Iron	5%	40%
Limestone	1%	67%
Magnesium	5%	40%
Metallic mineral	5%	40%
Nickel	12%	37%
Non-metallic minerals	3%	67%
Plaster	1%	67%
Platinum	5%	87%
Radioactive minerals	10%	63%
Salt	12%	60%
Silver	4%	87%

^aArticle 20 of Law 756 of 2002 (which modified Article 35 of Law 141 of 1994).

Notes: Royalty tax pay by each mineral and share of revenue allocated to the local municipality. Tabulated data from Law 756 of 2002 (and Law 141 of 1994, which was partially modified by Law 756 of 2002).

Based on the people's ownership rights of the country's non-renewable resources, revenues from such resources are to be distributed among all the population. The government presents a proposal to redesign the constitutional framework that governs the distribution of royalties in Colombia, particularly articles 360 and 361 of the Constitution, replacing the current system with the Royalties General System, to employ these resources (which are expected to grow in the coming years) to promote economic

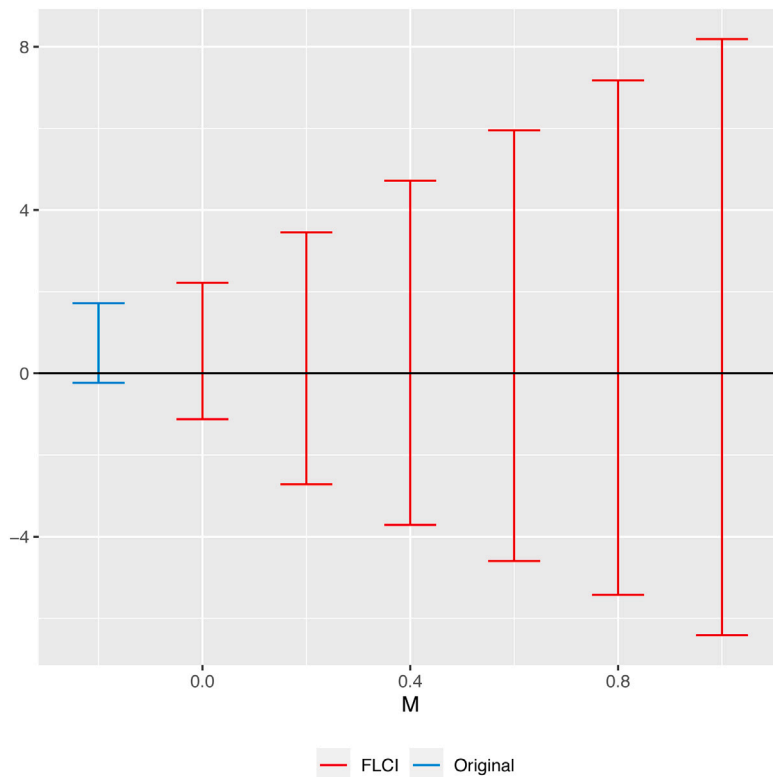


Fig. A.6. Robustness of the difference-in-difference estimates of Prediction 2 to violations of the parallel trends assumption. *Notes:* This figure examines how the robust confidence sets for the difference-in-difference estimates in Table 7 change as the parameter M (in $\Delta(M)$) varies following Rambachan and Roth (2020). M measures the change in slope of the differential trend. Typical difference-in-difference estimates assume $M = 0$ (i.e., that there is no differential trend in the absence of treatment). The mean change in slope in the pre-period in our data is 0.67. Given the large confidence intervals in Fig. 4(b), it is unsurprising our results are no longer statistically significant for small values of M . However, since both identification strategies support each other in the predictions of the model, we are not too concern by how sensitive our results are to the choice of M .

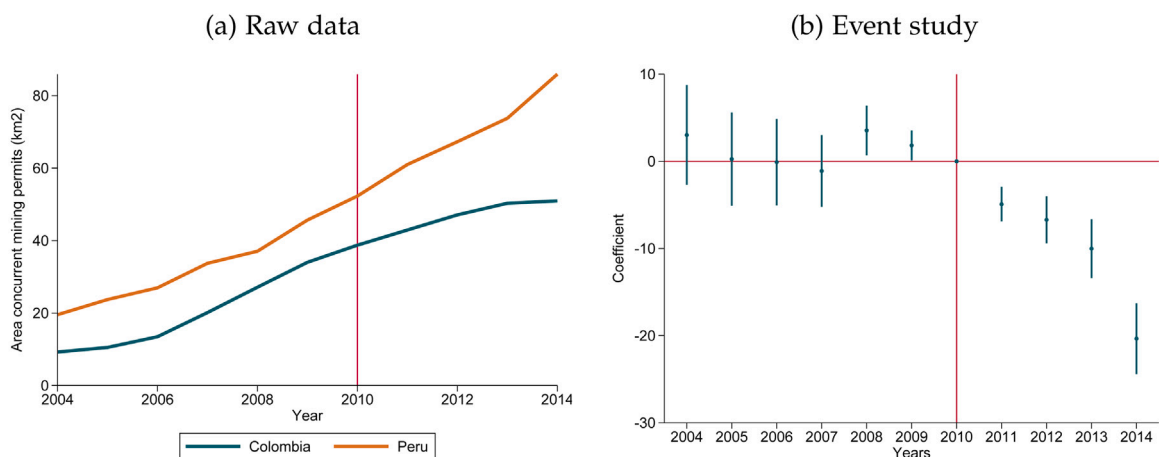


Fig. A.7. Parallel trends between Colombia and Peru in the area (km^2) of mining permits . *Notes:* Both panels use municipality-year data from 2004 to 2014 for Colombia and Peru. Fig. A.7(a) shows the evolution of the area (km^2) of mining permits in each country from 2004 to 2014. The estimates in Fig. A.7(b) are from an event study regression for the area (km^2) of mining permits. The x -axis plots time in years and the y -axis the coefficient of the indicator of Colombia interacted with the respective year. Point estimates and 95% confidence intervals are plotted. 2010 is the base year.

activity and equality among the countries' regions, fight poverty and increase competitiveness. Principles of macroeconomic saving and regional and social equality shall guide the new design of the royalties system”.

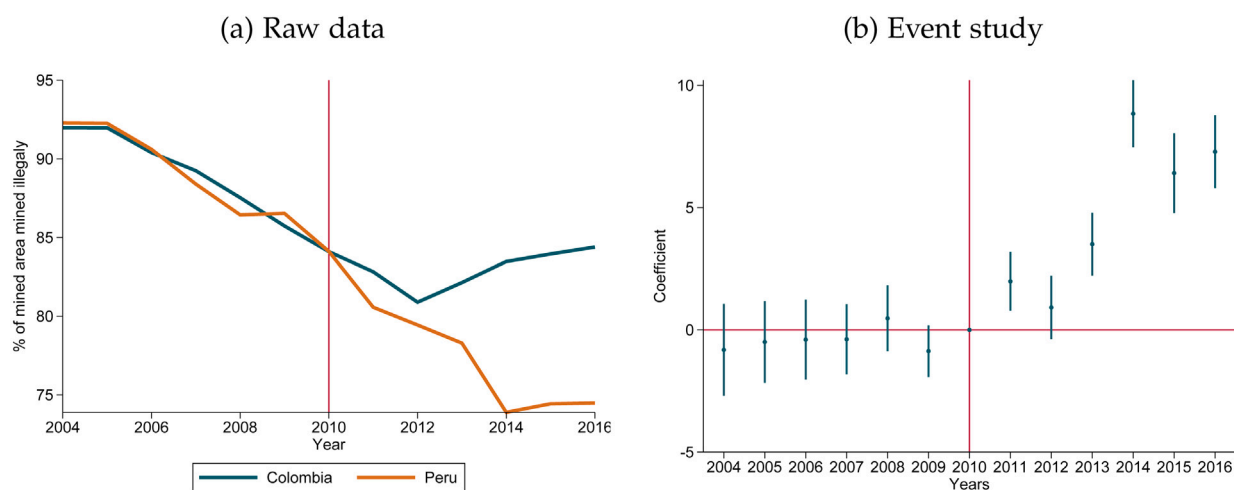


Fig. A.8. Percentage of mined area mined illegally in Colombia and Peru, including 2015 and 2016. *Notes:* Both panels use municipality-year data from 2004 to 2016 for Colombia and Peru. Fig. A.8(a) shows the evolution of the average percentage of mined area mined illegally in each country from 2004 to 2016. The estimates in Fig. A.8 are from an event study regression for the percentage of mined area mined illegally. The x -axis plots time in years and the y -axis the coefficient of the indicator of Colombia interacted with the respective year. Point estimates and 95% confidence intervals are plotted. 2010 is the base year. Since we lack information from mining titles emitted after 2014 in Peru, we assess legality in 2015 and 2016 using 2014 titles in both countries.

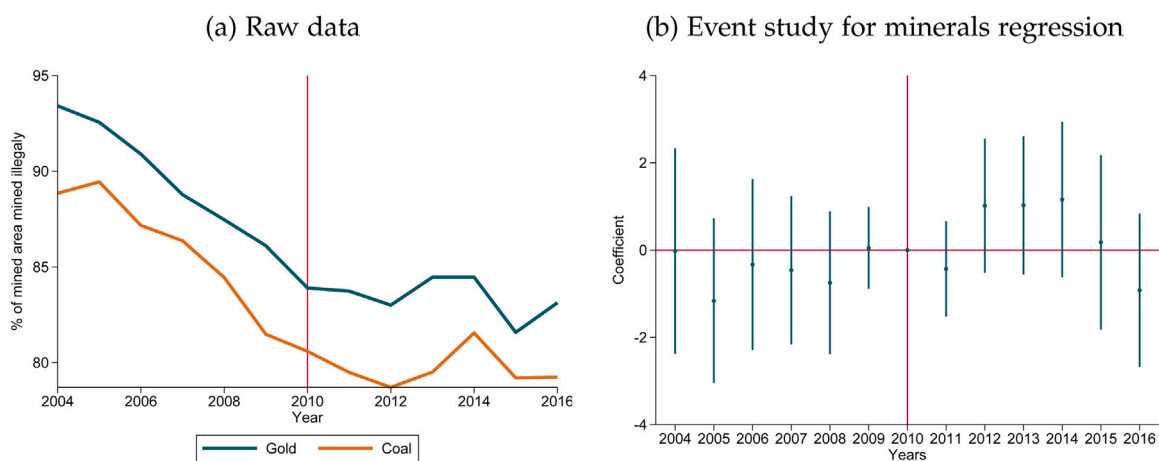


Fig. A.9. Percentage of mined area mined illegally in Colombia (Gold vs Coal), including 2015 and 2016. *Notes:* Both panels use municipality-mineral-year data from 2004 to 2016 for Colombian municipalities. Fig. A.9(a) shows the evolution of the average percentage of mined area mined illegally across municipalities for gold and coal mining in each year. The estimates in Fig. A.9(b) are from an event study regression for the percentage of mined area mined illegally by mineral. The x -axis plots time in years and the y -axis the coefficient of the royalty rate interacted with the respective year. 2010 is the base year. Since we lack information from mining titles emitted after 2014 in Peru, we assess legality in 2015 and 2016 using 2014 titles in both countries. Point estimates and 95% confidence intervals are plotted.

The constitutional reform bill (eventually known as Legislative Act 5 of 2011) was sent to Congress on August 31, 2010 and approved on July 18 of that year. The law regulating the new system (Law 1530 of 2012) was sent to Congress on October 14, 2011 and approved on May 17, 2012. Hence the new system went into effect after May 2012.

B.2. Suspension of new mining permit issuance

Between February 2, 2011 and December 31, 2012, the Colombian government did not accept requests for new mining permits. This was a consequence of Law 1382 of 2010, which mandated an upgrade of the platform used to grant and archive permits (Ministerio de Minas y Energía, 2011). However, we believe this closure of the “mining counter” (as it is known in Spanish: “ventanilla minera”) is unrelated to the immediate increase in illegal mining.

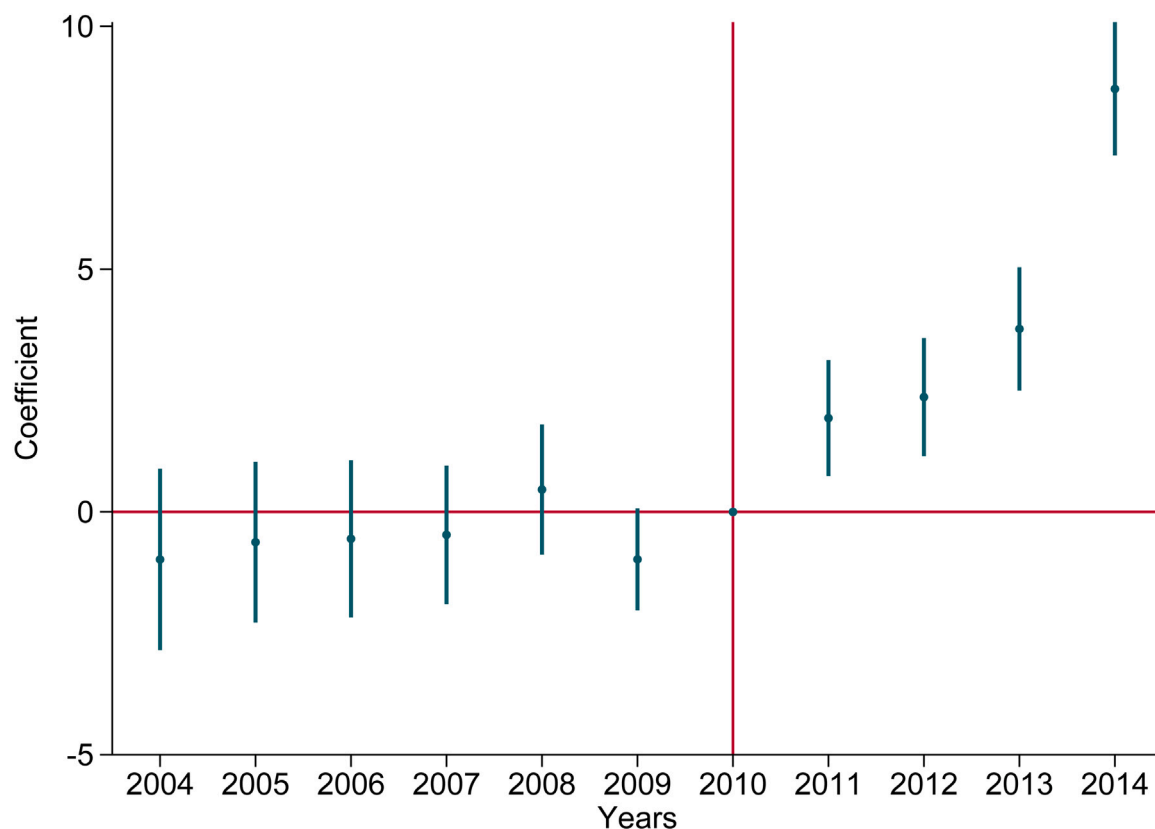


Fig. A.10. Event study Colombia vs. Peru: concurrent permits. *Notes:* The estimates are from an event study regression for the percentage of the mined area mined illegally using concurrent permits. This is analogous to Fig. 3, but using the permits granted each respective year. The x -axis plots time in years and the y -axis the coefficient of the indicator of Colombia interacted with the respective year. Point estimates and 95% confidence intervals are plotted. 2010 is the excluded year. The figures differ on whether for 2012 and 2013 we use the permits granted those years or the permits of 2014 when they reopened the system to register permits. Before the reform, there are no statistically significant differences between the two countries. After the reform, we observe a differential increase in Colombia.

Table A.2

Change in armed groups homicide rate.

Dependent variable: Armed group homicide rate	All	No AG Bef Reform	AG Bef Reform
	(1)	(2)	(3)
After X Col X % Budget Loss	-0.093 (0.19)	-0.057 (0.046)	0.0016 (0.38)
Mineral price index	0.12 (0.11)	0.0079 (0.022)	0.26 (0.29)
Time FE	Yes	Yes	Yes
N. of obs.	10,204	6,171	4,033
Municipalities	940	568	372
Mean of dep. var.	24.4	1.65	59.3
R ²	0.24	0.11	0.23

Notes: The entries in Table A.2 are the estimated coefficients of Eq. (4). The dependent variable is the percentage of the cumulative area mined that is mined illegally in the municipality. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The central government mining agency has been slow to respond to permit requests since before the suspension. The number of requests has exceeded institutional capacity since the early 2000s.⁵⁰ As of February 2, 2011, there were 19,629 pending requests (they had not been granted or denied). Of those, about 86% were filed between 2006 and 2010 (and the rest before 2006). As

⁵⁰ This is reflected as well in the proportion of “fiscalization” visits — inspection in which the agency verifies that mining companies are complying with regulation — which was around 60% in 2005 and 70% in 2010.

Table A.3

Production of mineral commodities in 2013.

Source: USGS <http://minerals.usgs.gov/minerals/pubs/country/sa.html>.

Country	Alum	Coal	Copper	Gold	Iron ore	Steel	Nickel	Silver	Tin
Brazil	34,171	7,407	271	79,573	386,270	34,163	105	–	16,830
Colombia	–	103,885	1	55,745	710	1,297	70	14	–
Ecuador	–	–	–	2,800	–	562	–	1	–
Panama	–	–	–	2,099	–	–	–	–	–
Peru	–	189	1,286	151,486	10,126	1,069	–	3,407	23,688
Venezuela	2,312	1,083	–	1,691	10,583	2,250	6	–	–

Notes: Alum stands for aluminum. Gold production in kilograms. Silver and tin production in metric tons. Other minerals in thousands of metric tons.

Table A.4

Summary statistics for municipalities separated by censused status.

	All	Censused	Not Censused	Difference
% Loss	–4.03 (11.6)	–5.14 (10.3)	–3.10 (12.5)	2.04*** (0.76)
Produced precious metals	0.29 (0.45)	0.31 (0.46)	0.27 (0.44)	–0.042 (0.030)
Oil production	0.080 (0.28)	0.040 (0.20)	0.12 (0.33)	0.080*** (0.018)
Armed group presence before reform	0.40 (0.49)	0.39 (0.49)	0.40 (0.49)	0.0074 (0.032)
Population	25280.0 (40628.4)	23160.5 (41049.0)	27072.4 (40223.3)	3911.9 (2685.3)
Area (km2)	638.1 (1330.7)	633.1 (1348.7)	642.4 (1316.7)	9.30 (88.1)

Notes: Summary statistics for municipalities, disaggregated by whether the municipality was included in the 2010 Mining Census. An observation is a municipality. All data comes from CEDE's municipalities panel. Calculations: Authors.

Table A.5

Mineral mined according to the mining Census.

Mineral	# of Mines	Percentage
Gold	3734	27.25
Coal	2778	20.28
Clay	2240	16.35
Sand	1994	14.55
Gravel	1046	7.63
Cement limestone	461	3.36
Silver	302	2.20
Esmeralds	268	1.96
Other	253	1.85
Salt	220	1.61
Rocks	220	1.61
Platinum	111	0.81
Magnesium silicate	74	0.54
Total	13701	100.00

Notes: Number of mines by mineral extracted according to the 2010 Mining Census.

of 2010, the average number of days before the agency replied to a request for a permit was 514 (Unidad de Planeación Minero Energética, 2014). During the period in which the “mining counter” was closed, the agency sought to finalize many of the pending requests and was presumably able to process requests more quickly once it reopened (although we were unable to find concrete data on this).

Given (a) the time lag between a request and a permit, and (b) the fact that the agency was presumably able to reply more promptly to requests after 2012, we do not believe this change would show up in the data in 2011. While all minerals were subject to the closure of the mining counter, the identification strategy, which relies on variation across minerals (and municipalities that can mine these minerals) to identify the effect of the reform, also shows an increase in illegal mining after the reform.

Appendix C. Theoretical framework

C.1. The decision to operate legally

Consider a miner with capital K who must decide whether to operate legally. If he operates legally (L), he has to pay the associated royalties α and permit fees $T(\text{Area}(K))$ to the central government. But if he decides to operate illegally (I) he makes a

Table A.6
Effect of the reform on illegal mining.

Dependent variable:	% mined area mined illegally		Area mining permits (km ²)
	Cumulative (1)	New (2)	Cumulative (3)
Panel A: Main results			
After x Colombia	4.21*** (0.60)	5.12*** (0.67)	-11.8*** (2.80)
N. of obs.	26,323	10,679	26,323
Panel B: Treatment starting in 2012			
After x Colombia	4.47*** (0.59)	5.35*** (0.70)	-12.6*** (2.58)
N. of obs.	26,323	11,592	30,021
Panel C: Sample without year 2011			
After x Colombia	4.78*** (0.64)	5.98*** (0.74)	-13.4*** (2.83)
N. of obs.	23,921	10,209	27,291
Panel D: Sample restricted to $year \in [2007, 2014]$			
After x Colombia	4.05*** (0.53)	4.84*** (0.62)	-11.4*** (2.03)
N. of obs.	19,520	8,664	21,799
Panel E: Sample restricted to $year \in [2004, 2010] \cup \{2014\}$			
After x Colombia	9.01*** (0.74)	6.72*** (0.99)	-21.6*** (3.18)
N. of obs.	18,719	8,165	21,797
Panel F: Full sample with municipality specific trends			
After x Colombia	1.85*** (0.64)	2.43*** (0.80)	-7.26*** (1.34)
N. of obs.	26,323	11,592	30,021

Notes: Estimated coefficients using Eq. (4) with different dependent variables. The estimations use municipality-year data from 2004 to 2014 for Colombia and Peru. In the first column, the dependent variable is the percentage of the cumulative area mined that is mined illegally in the municipality. Hence, Column 1 only includes municipalities with detected mining. In Column 2, we calculate this percentage only in the new area mined that year. The number of observations is smaller since the estimation requires a cloud-free image in consecutive years; also, we cannot use observations from 2004 because we do not have satellite images from the previous year. In the last column, the dependent variable is the total area of mining permits, measured in square kilometers. Panel A replicates the results from Table 4, using the whole sample and having the treatment start in 2011. Panel B uses the whole sample and but the treatment starts in 2012. Panel C removes the year 2011 from the sample, and the treatment starts in 2012. Panel D restricted the sample to years between 2007 and 2014 to have the same number of pre-treatment and post-treatment periods. Panel E restricted the sample to years between 2007 and 2010, in addition to 2014. Panel F uses the full sample and includes municipality specific trends. All regressions include municipality and time fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7
Results of selecting optimal controls with Double Lasso procedure.

Dependent variable:	% of mined area mined illegally		
	(1)	(2)	(3)
After x Colombia	5.09*** (0.59)	4.98*** (0.65)	5.46*** (0.64)
Controls	Main	All	DLasso
N. of obs.	22,459	22,459	22,459
Municipalities	2,517	2,517	2,517
Mean of Dep. Var.	83.84	83.84	83.84
R ²	0.75	0.75	0.75
Controls	Main	All	DLasso

Notes: The entries in Table A.7 are the estimated coefficients of Eq. (4). The dependent variable is the percentage of the cumulative area mined that is mined illegally in the municipality. "Main" repeats the estimates from the main differences-in-differences specification (Column 1). The number of observations is different from Table 4 because when lagged variables are included we lose the first year in the sample and we do not have controls for all municipalities. "All" includes the price index, population, armed group homicides, and all these variables squared, lagged, interacted among them, interacted with a linear trend, and interacted with the quadratic trend. "DLasso" includes the variables from "All" selected from a Double Lasso procedure: In this case, the model selects homicides, price, lagged price, lagged price squared, lagged homicides and price interacted with the population. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8

Robustness of the results to different specifications.

Dependent variable:	% of mined area mined illegally						
	All	Without Nat Parks	Prob	Border	Cutoff	Adjusted	Weights
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After x Colombia	4.21*** (0.60)	4.43*** (0.61)	4.66*** (0.41)	4.12*** (0.61)	4.79*** (0.41)	6.52*** (0.70)	5.58*** (0.63)
After x Col x Far				0.97 (1.63)			
N. of obs.	26,323	25,910	28,813	26,323	28,959	17,594	26,139
Municipalities	2,701	2,671	2,748	2,701	2,748	2,018	2,697
Mean of dep. var.	85.1	84.7	88.5	85.1	88.2	83.7	85.0
R ²	0.73	0.73	0.78	0.73	0.76	0.78	0.77

Notes: The entries in Table A.8 are the estimated coefficients of Eq. (4). The dependent variable is the percentage of the cumulative area mined that is mined illegally in the municipality. Column (1) is the main specification. Column (2) excludes mined areas in national parks, which cannot be legalized even if the permit fees were paid. Column (3) uses the probability that a pixel is mined instead of the dummy of mined. Column (4) includes an interaction with an indicator that the Colombian municipality is more than 1,000km from the border with Peru. Column (5) changes the cutoff of the mining predictions. It uses the cutoff where $TPR(\rho) - FPR(\rho)$ is maximized. Column (6) adjusts the measure of area mined according to Eq. (9). Column (7) weights each observation by the fraction of the municipality analyzed (i.e., cloud-free). All regressions include municipality and year fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9

Effect of the reform on illegal mining by mineral.

Dependent variable:	% mined area mined illegally		Area mining permits (km ²)
	Cumulative (1)	New (2)	Cumulative (3)
Panel A: Main results			
After $\times \alpha$	1.18** (0.58)	1.29* (0.76)	3.41 (2.19)
N. of obs.	6,774	4,918	6,774
Panel B: Treatment starting in 2012			
After $\times \alpha$	1.57** (0.64)	1.16 (0.86)	3.23 (1.98)
N. of obs.	6,774	4,918	6,774
Panel C: Sample without year 2011			
After $\times \alpha$	1.47** (0.67)	1.22 (0.88)	3.70 (2.27)
N. of obs.	6,088	4,394	6,088
Panel D: Sample restricted to $year \in [2007, 2014]$			
After $\times \alpha$	1.04* (0.57)	1.11 (0.87)	4.56*** (1.63)
N. of obs.	5,370	4,108	4,684
Panel E: Sample restricted to $year \in [2007, 2010] \cup \{2014\}$			
After $\times \alpha$	1.54* (0.83)	1.00 (1.17)	3.43 (2.34)
N. of obs.	4,512	3,174	4,512
Panel F: Full sample with municipality specific trends			
After $\times \alpha$	1.18** (0.60)	1.29 (0.80)	3.70 (2.35)
N. of obs.	6,774	4,918	6,088

Notes: The entries in Table 7 are the coefficients estimates of Eq. (5), where the dependent variable is the percentage of the area mined that is mined illegally in the municipality region with potential to mine a given mineral. The estimations use municipality-mineral-year data for Colombian municipalities from 2004 to 2014. Panel A replicates the results from Table 7, using the whole sample and having the treatment start in 2011. Panel B uses the whole sample and but the treatment starts in 2012. Panel C removes the year 2011 from the sample, and the treatment starts in 2012. Panel D restricted the sample to years between 2007 and 2014 to have the same number of pre-treatment and post-treatment periods. Panel E restricted the sample to years between 2007 and 2010, in addition to 2014. Panel F uses the full sample and includes municipality specific trends. All regressions include municipality-time and municipality-mineral fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10
Effect of the reform on illegal mining by mineral.

Dependent variable:	% mined area mineral mined illegally		
	(1)	(2)	(3)
After $\times \alpha$	1.18** (0.58)	0.91** (0.42)	0.51 (0.33)
N. of obs.	6,774	11,905	12,680
Municipalities	390	551	556
Mean of Dep. Var.	83.85	84.26	83.98
R^2	.96	.96	.95
Minerals:	Gold	Gold	Gold
	Coal	Coal	Coal
		Pt	Pt
			Mg

Notes: The entries in Table A.10 are the coefficients estimates of Eq. (5), where the dependent variable is the percentage of the area mined that is mined illegally in the municipality of a mineral. The first column includes municipality-time and municipality-mineral fixed effects and restricts the sample to gold and coal mines. The second column adds data for platinum (Pt) mines (the third most mined mineral according to Table A.5 for which we have data). The third column adds data for magnesium (Mg) (the next most mined mineral for which we have data). Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

side-payment $b(K)$ to the local authority⁵¹ and faces a probability of the illegal mine being detected by the National Police $Pr(K)$. This probability is increasing in the size of the mine.

The expected profits, in each case, can be expressed as:

$$\Pi_L = pq(K)(1 - \alpha) - C(q(K)) - T(\text{Area}(K))$$

$$\Pi_I = pq(K) - C(q(K)) - Pr(K)p_K K - b$$

where p is the international price of the mineral, $q(K)$ the quantity extracted as a function of K , α is the production tax paid by the firm, $C(\cdot)$ the associated cost of extraction, and p_K the price of capital. The cost of illegality is $Pr(K)p_K K$, because when an illegal mine is detected, its capital is confiscated or destroyed per the law (Section 2 provides more details). The side-payment is determined endogenously by each miner bargaining with the local authority depending on the payoffs for both when legal/illegal.⁵² We model the local authority as a single agent who values the budget of the municipality, the local externalities from mining, and the bribe he can obtain.⁵³ The local authority's payouts in each case are

$$G_L = f(pq\alpha\beta + B) - \gamma_L q$$

$$G_I = f(B) - \gamma_I q - Pr(K)V + b$$

where β is the share of royalty taxes allocated to the mining municipality, B is the municipality's budget aside from mining royalties, γ_i is the local environmental damage associated with each type of mining, and V is the cost to the local authority if the central government discovers the illegal mine and confirms the existence of collusion in a trial. This cost would be a monetary sanction or a prison sentence if evidence of the local authority receiving a bribe is found.⁵⁴ The function f reflects the valuation of the local municipality's budget by the local authority. We assume $f' > 0$, either because the local authority gets a share of the contracts or because it altruistically cares more about investing in local projects than in projects outside the municipality.

The "surplus" of illegal mining is the difference between the payoffs for the miner and the local authority when legal/illegal:

$$S(K) = \Pi_I - \Pi_L + G_I - G_L =$$

$$\underbrace{T + pq(K)\alpha}_{\text{Legality fees}} + \underbrace{f(B) - f(pq(K)\alpha\beta + B)}_{\text{Foregone revenue}} - \underbrace{Pr(K)(p_K K + V)}_{\text{Expected punishment}} - \underbrace{q(K)(\gamma_I - \gamma_L)}_{\text{Additional pollution}}$$

⁵¹ We are assuming the local authority observes all mining activity in its municipality without cost. Empirically this is supported by a survey of 18 local authorities, in which all confirmed that they were aware of the presence of illegal mining within their jurisdictions (Fedesarrollo, 2014). For a detailed case see Giraldo (2013) and <http://www.elpais.com.co/elpais/colombia/noticias/informe-exclusivo-denuncian-mafia-detras-mina-san-antonio-santander-quilichao>. Theoretically, in a model with endogenous effort to observe illegal mines the level of illegal mining is higher but the change in illegal mining with the reform is of similar magnitude.

⁵² The predictions on the surplus of illegal mining increasing do not require assumptions on the bargaining model. In Fig. C.1 in the Appendix we are assuming Nash bargaining with constant bargaining power before and after the reform.

⁵³ If the bribe was paid to an agent whose payoff does not depend on the municipal budget then the reform would not affect illegal mining under this framework.

⁵⁴ In most cases, the National Police destroys the machinery but does not conduct further investigation. Thus, we model V as zero.

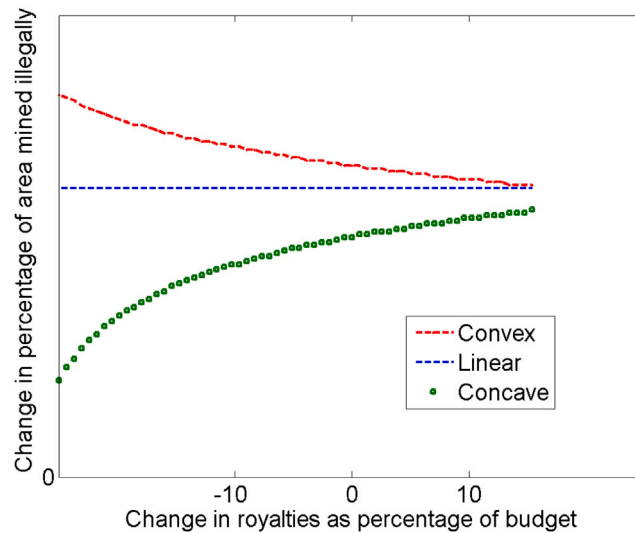


Fig. C.1. Theoretical predictions of the income effect of the reform. Change in percentage of area mined illegally before and after the reform, depending on the function the local authority uses to value the local municipality budget.

C.2. The income effect of the reform

The effect of the reform on the budget depends on the transfer/loss (B_1) based on socioeconomic criteria. The change in illegal mining surplus due to revenue lost with the reform can be written as:

$$\Delta S(K) = (f(B + B_1) - f(pq(K)\alpha\beta_1 + B + B_1)) - (f(B) - f(pq(K)\alpha\beta_0 + B))$$

The sign of $\Delta S(K)$ depends on the concavity of $f(\cdot)$. To see this, we separate $f(\cdot)$ into two components: $f(B) = \delta(B)B + g((1 - \delta(B))B)$, where the first term is the share of the budget that the local authority captures for itself and the second term is the valuation of the budget actually invested in public goods. If we assume the local authority captures a constant share of the budget $\delta(B) = \delta$ and $g(\cdot)$ is linear, then $f(\cdot)$ is linear. In that case, $\Delta S = pq\alpha(\beta_0 - \beta_1)$, which does not depend on B_1 , and the effect of the reform on illegal mining is the same for all municipalities regardless of whether they experience a net loss or win with the reform (i.e., there is no income effect). However, when the local authority has a convex function $f(\cdot)$, the surplus of illegal mining for any size K is now larger for municipalities whose budget decreased with the reform.⁵⁵ Consequently, the average size of illegal mines is larger for municipalities negatively affected by the reform and there should be a larger increase in illegal mining in these municipalities. The converse holds if the function is concave: The increase in illegal mining is larger in municipalities whose budget increased with the reform. Intuitively this is the case, when there is decreasing valuation to scale from royalty income. In short, the income effect of the reform depends on the concavity of f .

C.3. The optimal share of taxes for the local municipality

Using the increase in the area mined illegally with the reform, we make a first attempt to calculate the optimal share of taxes for the local municipality. Consider the problem of a central government that wants to distribute one dollar of possible tax revenue between mining and non-mining municipalities. Let β denote the share of taxes allocated to the mining municipalities. Let $e(\beta)$ be the function that relates evasion (e) to this share, so that only $1 - e(\beta)$ is available to be distributed among the municipalities. Given the model in Section 3 and the results, $e'(\cdot) < 0$: The greater the share allocated to the mining municipalities, the smaller the amount evaded. The objective is

$$\max_{\beta} (\lambda\beta + (1 - \beta))(1 - e(\beta)), \tag{7}$$

where λ is the weight of the mining municipality or how efficient a dollar spent in the mining municipality is, compared to a dollar spent in the nonmining municipality. If $\lambda \geq 1$, then it is optimal to allocate all the mining taxes to the mining municipality ($\beta = 1$), since evasion is minimized and spending is more efficient/preferred in these municipalities. If $\lambda < 1$, then:

$$\lambda = 1 + \frac{e'(\beta^*)}{1 - e(\beta^*) - \beta^*e'(\beta^*)} \tag{8}$$

⁵⁵ The same happens with a function with a reference point based on what the municipality received before the reform.

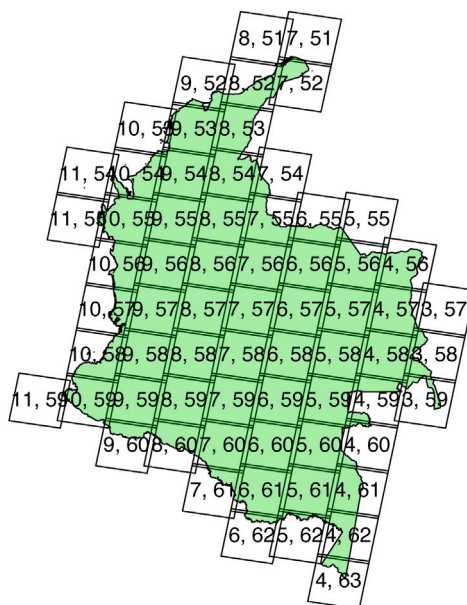


Fig. D.1. Scenes (Path, row) from Landsat -7 covering Colombia.

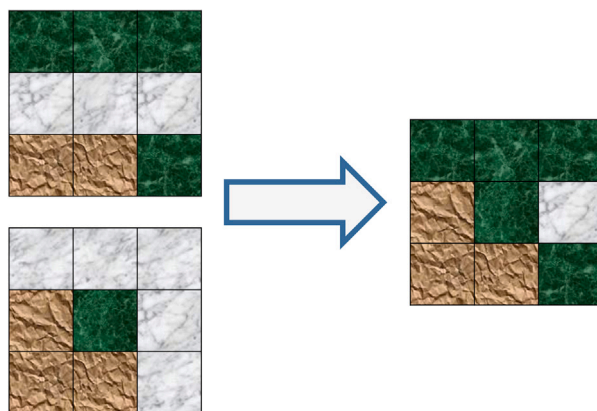


Fig. D.2. Creating a cloud-free mosaic.

We now substitute for the values estimated in the previous sections. After the reform we have the following: $\beta = 10\%$, $e(\beta) = 84\%$ (from Table 3), and $e'(\beta) = -0.11$ (from Table 4 divided by the change of β with the reform). Thus, $\lambda = 0.35$. That is, spending in nonmining municipalities needs to be three times as efficient as in mining municipalities for this to be optimal. However, Gallego et al. (2020) show that there are no differential effects per dollar spent in nonmining municipalities. Consequently, a larger share of taxes should be allocated to the mining municipalities, unless spending there is three times as preferred.

Appendix D. Constructing the illegal mining data

The combined area of Colombia and Peru is 2.42 million square kilometers. Hence, we analyze a total of 2.7×10^{10} pixels. In each pixel, we determine whether there is illegal mining. Below, we describe step by step the process we use. The full code to construct the illegal mining data can be found at https://dataverse.harvard.edu/dataverse/illegal_mining/ (see Fig. D.5).

1. We use images from the Landsat -7 satellite. The satellite takes a picture of each point on the Earth’s surface every two weeks. We use the web page of the U.S. Geological Survey <http://earthexplorer.usgs.gov/> to identify images that cover Colombia from 2004 o 2014. Fig. D.1 shows the identifiers for the images (scenes) that cover Colombia.
2. Download the necessary surface reflectance images from <http://espa.cr.usgs.gov/>. We request all images in UTM-18 projection. There are on average 550 images per year, each one around 230 MB (when compressed). A total of ~1.5TB of raw data.

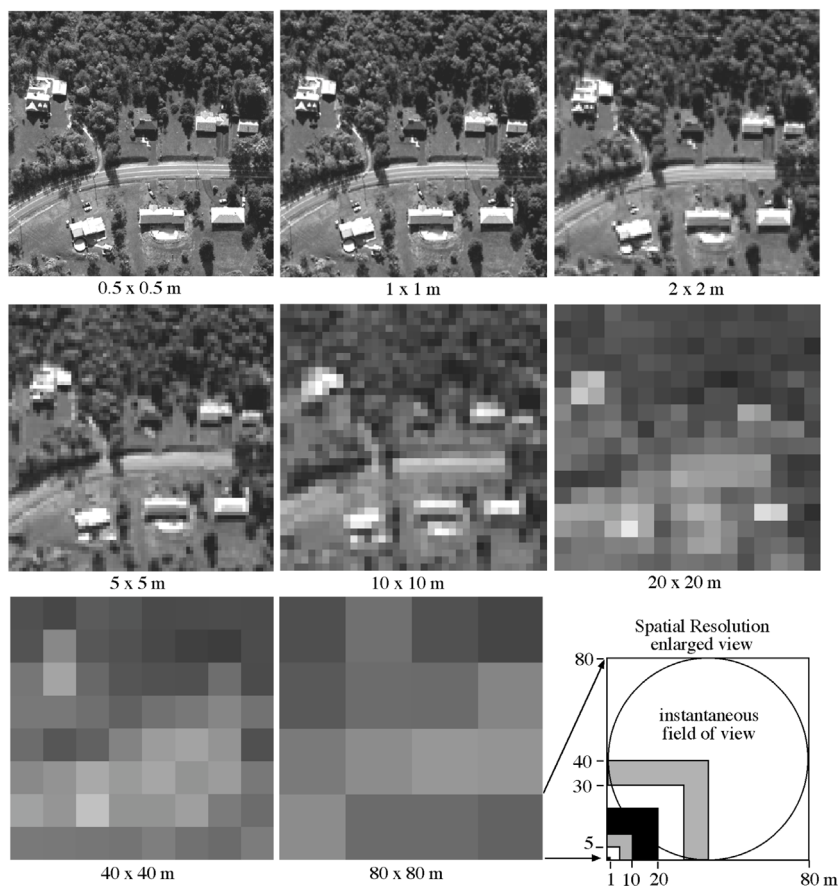


Fig. D.3. Illustration of satellite imagery of the same location with different resolutions. The imagery we use (Landsat) is 30×30 m resolution, so we cannot use shape recognition.

Source: Jensen (2007)

3. Given the presence of clouds, we need to construct a cloudless composite for every year. That is, we look for a cloudless image of each pixel and create a mosaic image with the cloud-free information from all images in a given year. Fig. D.2 presents a toy example of how we create a cloud-free mosaic. We use the R package *teamlucc* (<http://azvoleff.com/teamlucc.html>), with slight modifications, to remove clouds, adjust for topography, and create the mosaic. This process requires around 120 days of computing time.
4. The resolution of Landsat is 30×30 m so we cannot use shape recognition. See Fig. D.3 for an illustration. Instead, we use surface reflectance information to train a machine vector algorithm.
5. To train the prediction model we need to label pixels as mined or not mined. For this we use the 2010 Mining Census, which gives us the location and area of all the mines in half the municipalities of the country.⁵⁶ We only include open-pit mines, since those are the ones in which we expect to observe evidence of mining using the satellite images.
6. We confirm the presence of mines on the coordinates stated on the Census by using high-resolution images from Digital Globe (<https://www.digitalglobe.com/>). This allows us to draw the exact shape of the mine.
7. We also use the identified shape of mines in Open Street Map (<https://www.openstreetmap.org>) to complement the mining census.
8. Our training data frame consists of a matrix with 9 columns (variables) and 168,000 rows (observations or pixels). The columns are the 6 bands of the satellite information⁵⁷, the information on how long ago the pixel was deforested (Hansen et al., 2013), ecosystem type (Etter, 2006), and an indicator of whether the pixel is a mine or not (from the previous step). Fig. D.4 shows how the geospatial satellite information is transformed into a data frame.
9. We exclude from the analysis forested pixels using Hansen et al. (2013)'s deforestation data.

⁵⁶ Before using the Census data remove mines with coordinates outside the indicated municipality, or that have missing values, or that have nonsensical coordinates (e.g., minute or second values that are not between 0 and 60).

⁵⁷ Band 1 (blue), Band 2 (Green), Band 3 (Red), Band 4 (Near-infrared), Band 5 (Shortwave infrared 1) and Band 7 (Shortwave infrared 2).

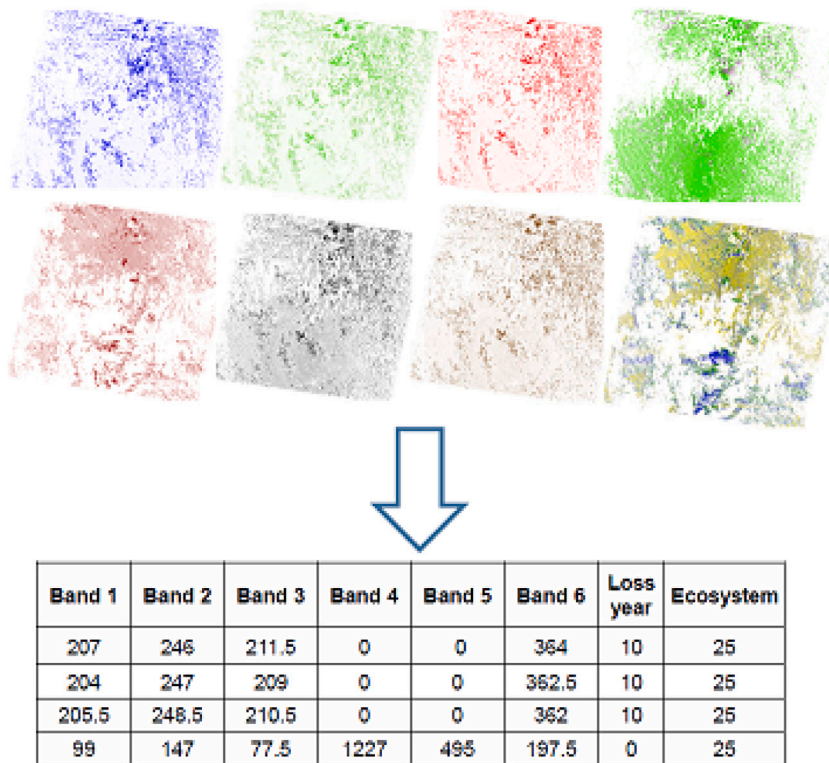


Fig. D.4. Visual representation of transforming the satellite data into a data frame.

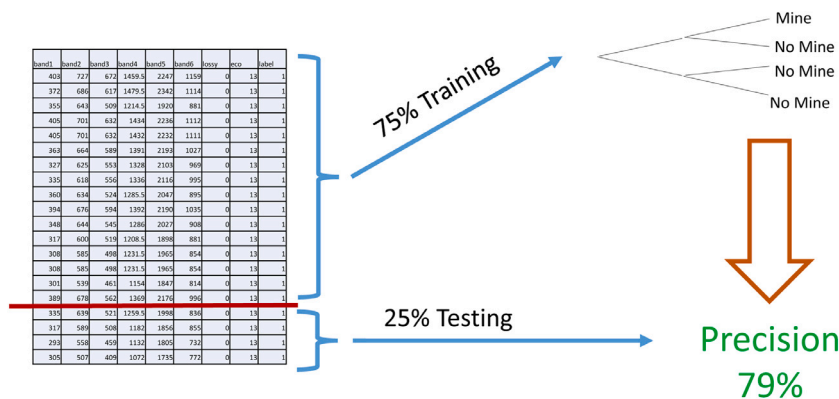


Fig. D.5. Visual representation of training and testing data.

10. We split the sample into training and testing sets, by dividing the country into $40km \times 40 km$ squares. We further subdivide each square into 4 squares and randomly choose one for testing and the other three for training. We do not take a random 25% sample for testing because each pixel is similar to its neighbors, so it is better to stratify this way.
11. We try boosting, support vector machines with radial kernels and random forest models in a small subsample of the data. For all three models we try down-sampling and SMOTE. We chose the best parameters for each case by 10-fold cross-validation. Based on the results in the subsample we decide to fit a random forest by down-sampling in the whole dataset.
12. The random forest consists of 100 trees so it is difficult to represent its structure. However, Fig. D.6 shows the relative “importance” (in what proportion of trees it appears) of each variable.
13. Once we have classified pixels as mined or not mined, the last step is to classify what mineral they are extracting. To determine the mineral being mined in each Colombian pixel, we use the map of mining potential produced by the National Mining Agency (Agencia Nacional de Minería, 2013). This map is based on research from the Colombian Geological Service, a scientific and technical institution in charge of determining the potential subsoil resources. If a mining pixel is located in

Variable importance in the mining prediction model

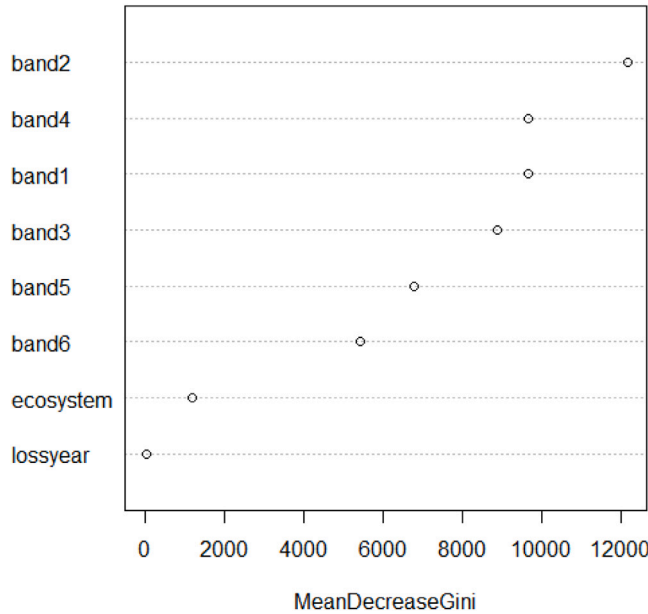


Fig. D.6. Importance of each variable for estimating whether a pixel is mined across all decision trees that make the random forest. band2, the green color band, is the one that helps distinguishes the most between mined and non-mined pixels. This makes sense because mined pixels lack vegetation.

an area with a single mineral (see Fig. D.7) we assign it that mineral. If a mining pixel is located in area with more than one potential mineral we resolve conflicts using the following priority rule: (1) Gold, (2) Platinum, (3) Copper, (4) Coal, (5) Columbite-Tantalite, and (6) all others alphabetically. For example, if a mining pixel is located in an area with Platinum and Coal, we assign it to Platinum.

D.1. Econometric analysis of the error term and implications for the optimal cutoff

It is important to analyze how the errors in the individual pixel prediction might affect our estimation of the effect of the reform on illegal mining. In this subsection, we explain how errors at the pixel level add to our measure of illegal mining area by municipality, and in turn how this might affect the coefficient estimates in the regression. Our estimated measure of mining area (\hat{y}_{mt}) in municipality m at time t can be expressed as the sum of correctly identified true mined pixels plus the misclassified true no-mined pixels:

$$\hat{y}_{mt} = \sum_{i \in Mines} (Pred(pix_i) = 1) + \sum_{i \notin Mines} (Pred(pix_i) = 1)$$

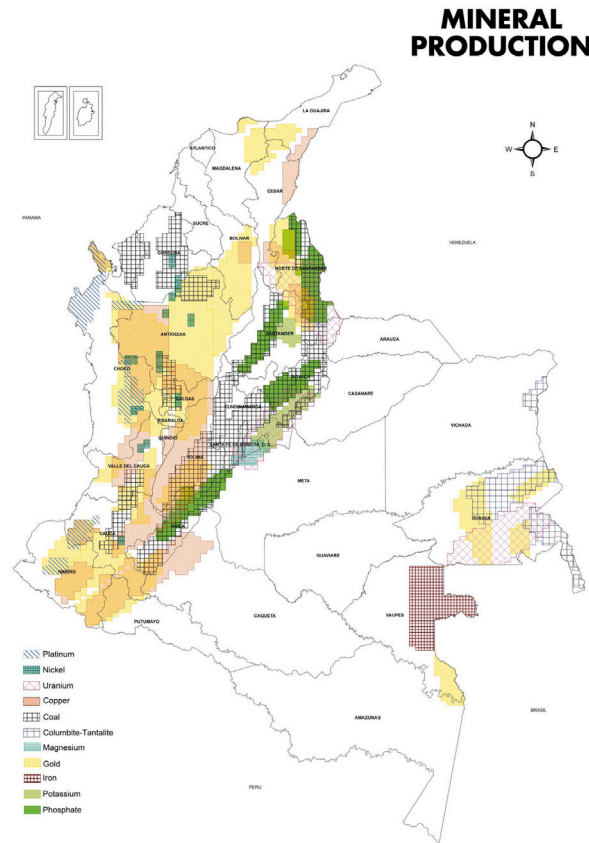
In each true mined pixel the probability of predicting a mine is equal to TPR and in each pixel that is truly mine-free the probability of predicting a mine is the FPR , where TPR and FPR are the true and false-positive rates of the prediction model respectively. In each pixel the random variable can be modeled as a Bernoulli, and, assuming independence and identical distribution, their sum is binomial.⁵⁸ As the number of pixels is large, we can approximate the sum with a normal. Thus $\hat{y}_{mt} = y_{mt}TPR + y_{Nmt}FPR + \epsilon_{mt}$, where y_{mt} is the true number of mined pixels, y_{Nmt} the true number of no-mine pixels and $\epsilon_{mt} \sim N(0, y_{mt}TPR(1 - TPR) + y_{Nmt}FPR(1 - FPR))$. Finally, since the total area of the municipality (Y_m) is fixed ($y_{Nmt} = Y_m - y_{mt}$) we can obtain the fraction of the municipality's area that is predicted to be mined as:

$$\frac{\hat{y}_{mt}}{Y_m} = \frac{y_{mt}}{Y_m} (TPR - FPR) + FPR + v_{mt} \tag{9}$$

where

$$v_{mt} \sim N \left(0, \frac{y_{mt}TPR(1 - TPR) + y_{Nmt}FPR(1 - FPR)}{Y_m^2} \right)$$

⁵⁸ We do not need to assume independence to prove a weaker version of the law of large numbers if we assume that the correlation between pixels far apart decays geometrically with distance. Appendix D.2 provides more details.



10

Fig. D.7. Mining potential.

Source: Reproduce from [Agencia Nacional de Minería \(2013, p.10\)](#).

The raw predicted fraction of the total municipality area that is mined underestimates the true fraction that is mined by a factor of $(TPR-FPR)$ plus an additive error term of FPR . Thus, the coefficient in the regression will underestimate the effect of the reform. When we use the predictions as the dependent variable in our regression analysis, a constant FPR will be absorbed by the municipality fixed effects.

To minimize the sum of squared errors, using formula (9), the optimal cutoff for declaring a pixel as mined is:

$$\rho^* = \operatorname{argmin}_{\rho} \sum_m \left(TPR(\rho) \frac{y_{m,2010}}{Y_{m,2010}} + FPR(\rho) \left(1 - \frac{y_{m,2010}}{Y_{m,2010}} \right) - \frac{y_{m,2010}}{Y_{m,2010}} \right)^2 \quad (10)$$

since 2010 is our training year from the mining Census. Since the fraction of the total municipality area that is mined is around 1%, the error of our predictions is approximately $1\%TPR + 99\%FPR$. This is why our cutoff (shown as the big dot in Fig. A.2) prioritizes having a small FPR . For completeness, in the results section, we present regressions with both the raw predictions and the adjusted predictions using formula (9).

D.2. Weak law of large numbers for correlated Bernoulli's random variables among pixels

The independence assumption is not necessary to prove a weaker version of the law of large numbers. Let us assume that $|cov(X_i, X_j)| \leq c^{dist(i,j)}$. We need to find a bound for $\sum_{j=1}^n cov(X_i, X_j)$. The largest sum of covariances will be for a pixel right in the center because it will be the shortest distance from the other pixels. For ease of exposition let us assume $n = (2k + 1)^2$, and consider pixel i in the center. This pixel will have its 8 neighbors, the 16 pixels surrounding them, and so on. The exact expression is:

$$\sum_{j=1}^n cov(X_i, X_j) \leq c + 8c^2 + 16c^3 + \dots + 8kc^{k+1}$$

With some manipulation, it can be shown that

$$\sum_{j=1}^n \text{cov}(X_i, X_j) \leq c + \frac{8c^2(1-c^k)k}{1-c}$$

Using Chebyshev's inequality we get the desired result.

References

- Acevedo, K.M., Bornacelly, I.D., 2014. Panel Municipal del CEDE. Documentos CEDE No. 012223, Universidad de los Andes-CEDE, Retrieved from <https://ideas.repec.org/p/col/000089/012223.html>.
- Agencia Nacional de Minería, 2013. Exploring opportunities: the answer is Colombia. Retrieved from <http://www.anm.gov.co/sites/default/files/Documentos/explorandopossibilidades.pdf>.
- Asner, G.P., Lactayo, W., Tupayachi, R., Luna, E.R., 2013. Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proc. Natl. Acad. Sci.* 110 (46), 18454–18459.
- Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci.* 113 (27), 7353–7360.
- Banerjee, A., Hanna, R., 2012. Corruption. In: Mullainathan, S., Gibbons, R., Roberts, J. (Eds.), *The Handbook of Organizational Economics*. In: *Introductory Chapters*, Princeton University Press.
- Baptiste, B., Pinedo-Vasquez, M., Gutierrez-Velez, V.H., Andrade, G.I., Vieira, P., Estupiñán-Suárez, L.M., Londoño, M.C., Laurance, W., Lee, T.M., 2017. Greening peace in Colombia. *Nat. Ecol. Evol.* 1 (4), 1–3.
- Belloni, A., Chernozhukov, V., Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. *J. Econ. Perspect.* 28 (2), 29–50.
- Brollo, F., Nannicini, T., Perotti, R., Tabellini, G., 2013. The political resource curse. *Am. Econ. Rev.* 103 (5), 1759–1796.
- Burgess, R., Hansen, M., Olken, B.A., Potapov, P., Sieber, S., 2012. The political economy of deforestation in the tropics. *Q. J. Econ.* 127 (4), 1707–1754.
- Cai, H., Treisman, D., 2004. State corroding federalism. *J. Public Econ.* 88 (3), 819–843.
- Cullen, J.B., Turner, N., Washington, E.L., 2021. Political alignment, attitudes toward government and tax evasion. *Am. Econ. J. Econ. Policy*.
- Echeverry, J.C., Alonso, G., García, A., 2011. Por qué es necesaria la creación de un Sistema General de Regalías. *Nota Fiscal No. 2*, Ministerio de Hacienda y Crédito Público, Retrieved from <https://ideas.repec.org/p/col/000089/012223.html>.
- El Espectador, 2017. Ofensiva de la Fiscalía contra la minería ilegal de oro. Retrieved from <https://www.elespectador.com/noticias/judicial/ofensiva-de-la-fiscalia-contra-la-mineria-ilegal-de-oro-articulo-685872>.
- Etter, A., 2006. Ecosistemas generales de Colombia. Downloaded from http://capacitacion.siac.ideam.gov.co/SIAC/home/Catalogo_mapas.html.
- Faber, B., Gaubert, C., 2019. Tourism and economic development: Evidence from Mexico's coastline. *Amer. Econ. Rev.* 109 (6), 2245–2293.
- Falkinger, J., et al., 1988. Tax evasion and equity: a theoretical analysis. *Public Finance Finances Publiques* 43 (3), 388–395.
- Fedesarrollo, 2014. Minería y medio ambiente en Colombia.
- Fisman, R., Wei, S.-J., 2004. Tax rates and tax evasion: Evidence from “missing imports” in China. *J. Polit. Econ.* 112 (2), 471–500.
- Fletcher, R., Saavedra, S., 2019. Job Migration in a Rivalry Setting. *Mimeo*.
- Foster, A., Gutierrez, E., Kumar, N., 2009. Voluntary compliance, pollution levels, and infant mortality in Mexico. *Amer. Econ. Rev.* 99 (2), 191–197.
- Gadenne, L., 2017. Tax me, but spend wisely? Sources of public finance and government accountability. *Am. Econ. J. Appl. Econ.* 9 (1), 274–314.
- Gadenne, L., Singhal, M., 2014. Decentralization in developing economies. *Annu. Rev. Econ.* 6 (1), 581–604.
- Gallego, J., Maldonado, S., Trujillo, L., 2020. From curse to blessing? institutional reform and resource booms in Colombia. *J. Econ. Behav. Organ.* 178, 174–193.
- Giraldo, J., 2013. El Gobierno del Oro en el Bajo Cauca, una lectura Weberiana sobre la explotación aurífera aluvial no legal. *Economía Criminal y Poder Político*, Medellín: Universidad EAFIT.
- Guiteras, R., Jina, A., Mobarak, A.M., 2015. Satellites, self-reports, and submersion: Exposure to floods in Bangladesh. *Amer. Econ. Rev.* 105 (5), 232–236.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160), 850–853.
- Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. *Amer. Econ. Rev.* 102 (2), 994–1028.
- Hoffman, V., Jakiela, P., Kremer, M., Sheely, R., 2017. There is no place like home: Theory and evidence on decentralization and politician preferences. *Unpublished*.
- Idrobo, N., Mejía, D., Tribin, A., 2014. Illegal gold mining and violence in Colombia. *Peace Econ. Peace Sci. Public Policy* 20, 83–111.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2014. *An Introduction to Statistical Learning with Applications in R*. Springer.
- Jayachandran, S., 2009. Air quality and early-life mortality evidence from Indonesia's wildfires. *J. Hum. Resour.* 44 (4), 916–954.
- Jensen, J., 2007. *Remote Sensing of the Environment: an Earth Resource Perspective*. Pearson Prentice Hall.
- Khan, A.Q., Khwaja, A.I., Olken, B.A., 2015. Tax farming redux: Experimental evidence on performance pay for tax collectors. *Q. J. Econ.* 131 (1), 219–271.
- La Republica, 2019. Reform al sistema de regalías. Retrieved from <https://www.larepublica.co/analisis/anif-2941063/reforma-al-sistema-general-de-regalias-act-legislativo-no-05-de-2019-2958694?>
- Lipscomb, M., Mobarak, A.M., 2016. Decentralization and pollution spillovers: evidence from the re-drawing of county borders in Brazil. *Rev. Econom. Stud.* 84 (1), 464–502.
- Martinez, L.R., 2019. Sources of revenue and government performance: Evidence from Colombia. Available at SSRN 3273001.
- Massé, F., Billon, P.L., 2018. Gold mining in Colombia, post-war crime and the peace agreement with the FARC. *Third World Thematics TWQ J.* 3 (1), 116–134. <http://dx.doi.org/10.1080/23802014.2017.1362322>.
- Ministerio de Minas y Energía, 1995. Decreto 1747 DE 1995.
- Ministerio de Minas y Energía, 2003. *Glosario técnico minero*.
- Ministerio de Minas y Energía, 2011. Resolución No. 180099 del 02 de Febrero de 2011.
- Ministerio de Minas y Energía, 2016. *Política Minera de Colombia*.
- OECD, 2018. In: Junghun, K., Dougherty, S. (Eds.), *Fiscal Decentralisation and Inclusive Growth*. OECD Fiscal Federalism Studies.
- OECD, 2019a. In: Moisis, A. (Ed.), *Asymmetric Decentralisation: policy Implications in Colombia*. OECD Multi-Level Governance Studies.
- OECD, 2019b. In: Dorothée Allain-Dupré, I.C., Moisis, A. (Eds.), *Making Decentralisation Work: a Handbook for Policy-Makers*. OECD Multi-Level Governance Studies.
- Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., Burke, M., Lobell, D., Ermon, S., 2018. Infrastructure quality assessment in africa using satellite imagery and deep learning. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 616–625.
- Pardo, A., 2013. Vacíos que motivan la corrupción minera siguen vigentes. Retrieved from <https://www.razonpublica.com/index.php/econom-y-sociedad-temas-29/7164-vacios-que-motivan-la-corrupcion-minera-siguen-vigentes.html>.
- de Planeación, D.N., 2007. Actualización de la cartilla “Las Regalías en Colombia”.
- Rambachan, A., Roth, J., 2020. *An Honest Approach to Parallel Trends*. *Mimeo*.

- Rettberg, A., Ortiz-Riomalo, J.F., 2016. Golden opportunity, or a new twist on the resource–conflict relationship: Links between the drug trade and illegal gold mining in Colombia. *World Dev.* 84, 82–96.
- Romero, M., Saavedra, S., 2015. The effect of gold mining on the health of newborns.
- Semana, R., 2013. Oro y crimen. Edition April 1st 2013.
- Shapiro, J., van den Eynde, O., 2021. Fiscal incentives for conflict: Evidence from India's red corridor. *Rev. Econ. Stat.*
- Sánchez de la Sierra, R., 2020. On the origins of the state: Stationary bandits and taxation in eastern congo. *J. Polit. Econ.* 128 (1), 32–74.
- Slemrod, J., 2019. Tax compliance and enforcement. *J. Econ. Lit.* 57 (4), 904–954.
- Slemrod, J., Weber, C., 2012. Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *Int. Tax Public Finance* 19 (1), 25–53.
- Slemrod, J., Yitzhaki, S., 2002. Tax avoidance, evasion, and administration. In: *Handbook of Public Economics*, Vol. 3. Elsevier, pp. 1423–1470.
- Solarte, L., 2017. Redes de minería ilegal en el Cauca y revocatoria. Retrieved from <https://www.las2orillas.co/redes-mineria-ilegal-cauca-revocatoria/>.
- Swenson, J.J., Carter, C.E., Domec, J.-C., Delgado, C.I., 2011. Gold mining in the Peruvian Amazon: Global prices, deforestation, and mercury imports. *PLoS One* 6 (4), e18875.
- The Global Initiative Against Transnational Organized Crime, 2016. Organized crime and illegally mined gold in Latin America.
- Unidad de Planeación Minero Energética, 2014. Indicadores de Minería en Colombia, 2014.
- UNODC, United Nations Office on Drug and Crime, 2016. Colombia: Explotación de oro de aluvión. Evidencias a partir de percepción remota.
- World Bank, 2017. Total natural resources rents (% of GDP). data retrieved from World Development Indicators, <https://data.worldbank.org/indicator/NY.GDP.TOTL.RT.ZS?locations=XM>.
- Zhuravskaya, E.V., 2000. Incentives to provide local public goods: fiscal federalism, Russian style. *J. Public Econ.* 76 (3), 337–368. [http://dx.doi.org/10.1016/S0047-2727\(99\)00090-0](http://dx.doi.org/10.1016/S0047-2727(99)00090-0), Retrieved from <http://www.sciencedirect.com/science/article/pii/S0047272799000900>.