Universidad del Rosario

Facultad de Economía

Local Incentives and National Tax Evasion: Unintended Effects of a Mining Royalties Reform in Colombia

Mauricio Romero
Santiago Saavedra

SERIE DOCUMENTOS DE TRABAJO

No. 234

Octubre de 2019

LOCAL INCENTIVES AND NATIONAL TAX EVASION: UNINTENDED EFFECTS OF A MINING ROYALTIES REFORM IN COLOMBIA*

Santiago Saavedra† Mauricio Romero‡

July 11, 2019

Abstract

Achieving a fair distribution of resources is one of the key goals of fiscal policy. To do this, governments often transfer tax resources from rich to marginalized areas. We study whether lower transfers dampen the incentives of local authorities to curb tax evasion in the context of mining in Colombia. To overcome the challenge of measuring evasion, we use machine learning on satellite images. Using difference-in-differences strategies, we find that a reduction in the share of revenue transferred back to mining municipalities led to an increase in illegal mining. This result illustrates the difficulties of redistributing tax revenues.

JEL classification: H26,O13,O17

*We are grateful to Prashant Bharadwaj, Arun Chandrasekhar, Pascaline Dupas, Larry Goulder, and Melanie Morten for their invaluable guidance. This paper was enriched by comments from 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) and Policy Design and Evaluation Lab (PDEL). Saavedra acknowledges support of the Stanford Institute for Economic Policy Research through the Kohlhagen Fellowship Fund. Romero acknowledges financial support from the Asociación Mexicana de Cultura.

†Universidad del Rosario, Department of Economics, santiago.saavedrap@urosario.edu.co.

‡ITAM, Centro de Investigación Económica, mtromero@itam.mx.

1 Introduction

Achieving a fair distribution of resources is one of the key goals of fiscal policy. To do this, governments often transfer tax resources from rich to marginalized areas. A smaller share of revenue transferred back to the locality where taxed economic activity takes place could dampen the incentives of local authorities to curb tax evasion (Banerjee, Mullainathan, & Hanna, 2012). In this paper we study this dimension of tax evasion in the context of illegal mining in Colombia. We find that a reduction in the share of mining royalties transferred back to the mining municipalities led to an increase in illegal mining.

A challenge of studying tax evasion (or any illegal activity) is measuring its extent (Banerjee et al., 2012; Slemrod & Weber, 2012; Fisman & Wei, 2004). To overcome this obstacle, we construct a novel dataset using machine learning predictions on satellite-imagery features to detect illegal mining. We assess the legality of identified mining activity using georeferenced mining permits issued by the national government.[1] We approximate the mineral mined in each location, based on research done by the National Mining Agency on the potential subsoil resources in the country (Agencia Nacional Minera, 2013). We use our measurement of illegal mining to study the effect of a reform that sharply reduced the share of royalties tax revenue transferred back to the municipality where the mine is located.[2] As a result, this reform reduced the incentives of local authorities to monitor miners' compliance with national regulations.

Traditional models of tax evasion, which do not take into account where tax revenue is

[1] We measure illegal mining following the definition of Colombia's national government as "mining activity without a mining title registered with the National Mining Registry" (Ministerio de Minas y Energia, 2003, p. 108). Not holding a mining title is highly correlated with the evading of royalty taxes.

[2] The government distributes the remaining revenue among all municipalities according to socioeconomic indicators.

spent and in which enforcement is taken as exogenous, will predict a null effect of the reform (Slemrod, 2018). Thus, we present a simple theoretical framework—in which a miner decides whether or not to operate legally—to formalize our intuition on how the reform affects the incentives of local authorities. We assume the local authority observes mining activity in its municipality. Thus, to operate illegally, the miner has to 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. In the post reform period legal mining has a lower payout for the local authority, thus the authority has less incentive to enforce regulations, thereby shrinking the bribe needed to operate illegally.

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 national 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 larger mines have a higher probability of detection and a bigger penalty. With these assumptions, the model yields four predictions: 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 effect of the reform is larger in areas where the national government's presence is weak. Finally, the reform should not have an effect on illegal mining in municipalities where where illegal armed groups are the *de facto* local authorities.

We use three different identification strategies to test the predictions of our model. We use 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. For example, since gold has a higher royalty rate than coal, we expect a larger in-

crease in illegal mining in gold-mining areas compared to coal-mining areas, even within the same municipality. This strategy allows us to use municipality-time fixed effects that capture time-varying confounding events like new mayors taking office, the peace process with the FARC-EP guerrillas, and changes in the title request system in Colombia.[3] Across minerals, we find that for every percentage point of royalty tax rate, there was an increase of 1.46 percentage points in the area being mined illegally after the reform. This increase in illegal mining resulted in approximately USD 270 million of foregone revenue or 45% of the USD 594 million in mining royalties to be distributed in 2015.

To test that mines are more likely to operate illegally after the reform, we use a difference-in-differences strategy that compares illegal mining in Colombia and Peru, before and after the reform. Illegal mining, as a share of total mined area, increased in Colombia by 4.47 percentage points after the reform relative to the change in Peru. The magnitude of the effect is close to four times what we estimated with the previous identification strategy. This is consistent with the fact that royalty taxes are around four percentage points higher in Colombia than in Peru. While the treatment estimate is potentially confounded by other changes that may affect illegal mining in either of the two countries, we chose Peru as the control group to minimize this risk (Section 5.1 provides more details). In addition, the fact that the estimated treatment effects are consistent across both identification strategies suggests this is not the case. Besides an increase in illegal mining area, evasion through production underreporting could be a margin of adjustment. Yet we do not find an effect of the reform on the reported production of legal mines. Since it is more difficult for local authorities to observe production than mining area, it is unsurprising that the reform only affects the latter.

[3]This identification strategy is closely related to that of Sanchez de la Sierra (2017)—which studies mineral taxation in Congo DRC—but we exploit within-municipality variation instead of across municipalities.

Finally, we use cross-sectional variation in municipality characteristics to identify the heterogeneous effect of the reform on municipalities with weak national government presence (prediction 3) and illegal armed group presence (prediction 4). As predicted by the model, the effect of the reform is greater in municipalities where the national government's presence is weak. Likewise, illegal mining did not increase in municipalities where the power is *de facto* held by illegal armed groups.

To the best of our knowledge, this is the first paper to quantify the response of tax evasion to the share of tax revenue allocated to local spending.[4] This paper joins a large literature on the determinants of tax evasion (see Slemrod (2018) for a review). A closely related paper by Khan, Khwaja, and Olken (2015) presents experimental evidence that performance pay for tax collectors increased both tax revenue and reported bribes. Local authorities in our context are not direct tax collectors. However, we find an analogous effect when incentives for tax collection are reduced.

We also contribute to the small but burgeoning literature on the political economy of natural resources management. Like Burgess, Hansen, Olken, Potapov, and Sieber (2012) and Lipscomb and Mobarak (2016), we study a national interest resource, the monitoring monitoring of which depends on local authorities. Both Burgess et al. (2012), and Lipscomb and Mobarak (2016) find that political incentives of local government officials increased deforestation and water pollution. We complement their findings by showing that fiscal incentives are also an important determinant of environmental degradation.

Finally, we contribute to the public finance literature on how revenues should be collected and distributed among different levels of government (Gadenne & Singhal, 2014). Previous papers have shown that politicians politicians value spending on projects for

[4] Cai and Treisman (2004) provide examples of specific cases where regional governments helped firms evade national taxes and regulations. We provide causal evidence using quasi-experimental variation.

citizens in their home areas twice as much as general government spending (Hoffman, Jakiela, Kremer, & Sheely, 2017) and that tax revenue is more efficiently spent than grant revenue (Gadenne, 2017). Our paper, instead of focusing on spending, addresses how spending allocation affects tax evasion. In principle, our results could also be explained by a model of tax morale, where citizens evade more because the revenue is not spent on their municipality (Falkinger et al., 1988; Cullen, Turner, & Washington, 2018). However, we do not think this is the case in our setting, since less than 25% of mine owners are from the same municipality where the mine is located.

Methodologically, this paper uses machine learning both to construct the dependent variable and to estimate causal effects. We use applications of machine learning techniques for causal inference (Belloni, Chernozhukov, & Hansen, 2014; Athey & Imbens, 2016) and join the growing body of literature that uses satellite observations to study economic outcomes (e.g., Foster, Gutierrez, and Kumar (2009); Jayachandran (2009); Henderson, Storeygard, and Weil (2012); Guiteras, Jina, and Mobarak (2015); Faber and Gaubert (2019)). Previous papers studying illegal mining used static measures in their analysis (Idrobo, Mejia, & Tribin, 2014; Romero & Saavedra, 2015). Thus, our panel dataset on illegal mining by municipality is a contribution in itself, as is the algorithm used to create the dataset, which could be used to create similar datasets 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. 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, Llactayo, Tupayachi, & Luna, 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 national government owns subsoil resources, including minerals. This 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. Colombia's national government allocates mining permits and sets royalty taxes for mineral extraction. The title holder pays a fee equal to a daily minimum wage per hectare per year.[5] Additionally, mining companies pay royalties based on the gross value and type of minerals extracted.[6] Royalties vary across minerals and depend on the quantity extracted. For example, the royalty rate for construction materials is 1%, the rate 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 title in the neighboring country of Peru, broadly speaking, has the same implications as in Colombia. In both countries, all natural resources belong to the state and mining title holders must follow environmental regulations and pay associated royalties.[7] In Section 5.1 we discuss our reasons for choosing Peru as the control country.

[5] If the title area is between 2,000 and 5,000 hectares, the title holder pays two times the legal minimum wage per hectare, while holders of titles to areas larger than 5,000 hectares pay three times the minimum wage per hectare (Agencia Nacional Minera, 2013).

[6] 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).

[7] See Ministerio de Minas y Energia (2016) for more details.

2.2 The reform

Before 2012, a mining municipality in Colombia would receive 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, while 40% are earmarked for regional funds and the rest must be saved.[8] The Congress of Colombia 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 budgets of municipalities (Table 1, Panel A). But this masks heterogeneity by winners and losers in the redistribution process.

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.[9] Illegal mining was not mentioned as a motivation for the reform, nor were the impacts of the reform on illegal mining contemplated.[10]

The reform did not alter the broad categories for which royalty revenue could be used. Over 90% of royalty revenue must be used to meet unsatisfied basic needs related to health, education, water supply and sewage treatment. After specific goals are achieved, 75% of the revenue must be allocated for these purposes and the rest can be used for environmental conservation and restoration, other education and health projects, or other public works (Ministerio de Minas y Energia, 1995; Departamento Nacional de

[8]The monies allocated to the regional funds are distributed according to population and poverty and unemployment rates; thus, the net impact of the reform in each municipality varies according to these characteristics. 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.

[9]See https://www.sgr.gov.co/LinkClick.aspx?fileticket=bsf8qrvGVOg%3D&tabid=181

[10]Congress approved the reform six months before it was implemented. Thus, we cannot rule out some anticipation by local governments of its effects. However, we argue the timing of the reform is exogenous to the evolution of illegal mining. Appendix B.1 provides a detailed time line of the reform.

Planeación, 2007).[11]

2.3 Illegal mining

Illegal mining is widespread in Colombia. In 2010, the national government conducted a census of mines, including those without a title, in half of the country's municipalities. According to the census, 36% of the reported mined area was not under a title. The United Nations Office on Drug and Crime (UNODC) manually inspected satellite images from 2014 and found that 81% of the area mined for gold was illegally mined (UNODC, 2016). The UNODC (2016) estimate of overall illegal mining is comparable to the 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 is consistent with some illegal mine operators refusing to answer the census or underreporting their mining area. We do not use the census directly for our estimates.

There have been three attempts to legalize illegal mines in Colombia, all with little success.[12] Government incentives for the legalization of illegal mines 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 title.[13]

Local authorities in Colombia are responsible for enforcing mining laws in their areas.

[11]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%.

[12]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 one title legalization was approved. Finally, a pilot legalization program that started with 150 mining operations in 2012 only has 25 still in the process after three years, and none have met all the requirements (The Global Initiative Against Transnational Organized Crime, 2016).

[13]Before the decree was issued, the machinery was supposed to be confiscated; this was difficult to implement in remote regions.

This includes suspending any mining activity carried out without a title, according to Law 685 of 2001. In addition, they must confiscate any minerals without a certificate of origin from a legal mine. If the suspension was ineffective, local authorities must inform national law enforcement authorities that they should proceed with confiscation/destruction of machinery.

A noteworthy feature of Colombia is that both local police forces and the national army are funded by the central government. Thus, the reform did not alter the funding mechanism for local enforcement agencies.

3 Theoretical framework

We present a framework for understanding a miner's decision to operate legally or illegally. This decision depends on the fees imposed on legal operations and the assumptions on penalty and probability of detection of illegal mines. We derive four theoretical predictions based on reasonable assumptions.

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:[14]

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

[14] Appendix C provides a detailed derivation of this expression.

where T are the title 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. β is the share of royalties allocated to the mining municipality— —the parameter affected by the reform studied on this paper.[15] $Pr(K)$ is the probability that the national government independently detects the illegal mine, and $\Theta(K)$ the penalty for the miner and the local authority if the mine is detected.

Any firm with capital K such that $S(K) \geq 0$ will pay the bribe and operate illegally.[16] We denote by K^* the value of capital such that $S(K^*) = 0$. Any firm with $K > K^*$ will operate legally, assuming returns to capital do not increase faster than expected punishment of capital destruction if caught operating illegally.

3.2 The effect of the reform on illegal mining

The reform did not change the fees paid to operate legally; nor did it change the royalty rate (α) paid by each mining firm. 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. As seen in Figure 1, the associated effect 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 with increasing probability of detection and punishment, a smaller β increases illegal mining, since the surplus of illegal mining is larger for every level of

[15] We could add a function f that reflects the valuation of the local municipality's budget by the local authority. We assume $f' > 0$, which can be justified by either the local authority getting a share of the revenue from the contracts or because it cares more about investing in local projects than in projects outside the municipality. The shape of f will play an important role when studying the income effect of the reform (Appendix C provides more details).

[16] The model assumes K is exogenous and follows a uniform distribution. In addition, firms are unable to merge.

capital. In particular, the threshold size of illegal mines (K^*) increases and the share of area mined illegally increases. To see this, note that the share of mined area mined illegally is:

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

Empirically larger mines are easier to observe, and given that Colombian law dictates confiscation/destruction of machinery, the penalty for larger mines is larger. As shown in Figure 1, under these assumptions, after the reform the share of area that is mined illegally increases. This leads to the following prediction:

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

The decision model applies not only to a miner choosing to open a new mine, but also to a miner choosing to operate legally, if currently he is not, in each period. In the empirical section, we will test this hypothesis both in the stock of mined area and the new area mined each year.

3.3 Heterogeneous effects of the reform

Consider the change in the surplus of illegal mining (Equation 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 area illegally mined should be larger for minerals with a higher royalty rate.

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

Consider now a municipality where the probability of detecting illegal mines is small, because of weak presence of the national government ($Pr_{WP}() < Pr()$). 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. If the capital distribution is smooth at the threshold, we expect the reform to have a larger effect on the share of area mined illegally in municipalities with low probability of detection.[17]

Prediction 3 *The increase in the share of area mined illegally is larger in municipalities with 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 can apply in Colombian municipalities in areas where the local power is de facto held by illegal armed groups.[18] In a less extreme case, the effect of the reform is not null but smaller.

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

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

[18] 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).

This simple model has some limitations. First, it does not consider the location decision of the miner, given that mineral resources are fixed in the subsoil. However, a miner could move to a neighboring municipality, where conditions are more favorable à la Burgess et al. (2012). Second, we abstract from income effects when a local authority receives bribes from many miners. These two limitations affect the level and location of illegal activity, but not the qualitative effect of the reform. Finally, we model the decision to get a legal title and pay royalties as a single choice, but some legal mines may evade a percentage of production taxes. To explore this possibility, we assess the effect of the reform on the produced quantity reported by legal mines.

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; details are provided in the next subsection. The second database is from Colombia's national government mineral information system, SIMCO, on reported production.[19] We also use the municipality panel from the Center for Economic Development Studies (CEDE) at Universidad de los Andes with information on royalties, municipal budgets, institutional presence of the national government, armed groups presence, and other characteristics of Colombian municipalities (Acevedo & Bornacelly, 2014). Finally, for Peru we constructed a panel with analogous characteristics for each municipality.[20]

We present summary statistics for municipalities in Colombia and Peru in Table 1 Panels B and C, respectively. Municipalities in each country had a similar amount of mined

[19] http://www.simco.gov.co/

[20] 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).

area, although there are more inhabitants in Colombian municipalities and more area is titled in Peru. We exclude from the analysis municipalities without mining potential in the subsoil, because, tautologically, there can be no mining in those municipalities.

4.1 Constructing the illegal mining panel

We have information from the 2010 Colombian Mining Census on the locations of legal and illegal mines.[21] The census covered half of Colombian municipalities (Table 1 Panel D). Most of the mines are open pit. Hence, they can be observed from space 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 years. There are six main steps in the construction of 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 a certain 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 titles. (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 Minera, 2013). (vi) Collapse the results at the municipality-mineral level for the regression analysis.

In order to train the model, 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, deforestation year

[21] The mining census, published by Colombia's Ministry of Mines, surveyed only half of the country's municipalities. Although there may be concern that the municipalities sampled by the census were selected based on certain characteristics, we do not find any evidence of this (Table A.3 in the Appendix provides more details). Municipalities included and not included in the census are balanced in terms of change in royalties due to the reform, production of different minerals, institutional presence of the national government, and presence of armed groups.

(Hansen et al., 2013), and ecosystem type (Etter, 2006).[22] 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 Figure 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 with 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.[23] The aim is 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 that are suitable for this type of problem (James, Witten, Hastie, & Tibshirani, 2014). A random forest, as its name indicates, is a collection of many binary decision trees where in each node the candidate subset of explanatory variables to be used in the binary partition is random. Although we attempted to use a logit model, more familiar to many applied researchers, it generated twenty times more false positives.

The random forest prediction 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.[24] For each cutoff, we plot in Figure A.1 the associated true positive rate (TPR) and

[22]We use data from NASA's LANDSAT 7 satellite. Different wavelengths are captured 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. http://lpdaac.usgs.gov

[23]The name "tree" comes from the graphical representation of the nested binary decision rules.

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

false positive rate (FPR) in the testing sample. Ideally, we want to have 100% TPR and 0% FPR (upper left corner). As we lower the cutoff, we improve the TPR but also increase the FPR. In the literature it is standard to choose the cutoff ρ such that $TPR(\rho) - FPR(\rho)$ is maximized (marked with a blue star in the figure). There are two important aspects of our analysis and data that make this standard cutoff inappropriate. First, we are using the predictions as the dependent variable. Second, our sample includes many unmined pixels. We discuss both issues in subsection D.1 in the Appendix. In a nutshell, the formula we use to choose the optimal cutoff minimizes the sum of squared errors in the treatment effect regressions we estimate. The effect is that more weight is given to a lower FPR, given that most pixels in the country are unmined. Although we use this optimal cutoff in our main regressions, we also show that our results are robust to using the standard cutoff.

In Table 2 we present the confusion matrix for the optimally chosen cutoff. This 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.[25] 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). Appendix D provides further details on the steps to construct the illegal mining panel.

using the raw probability in Table 7 and Table A.7. The results are pretty similar.

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

4.2 From pixel predictions to municipality panel

After predicting whether a given pixel is mined in each year, we overlay the map of legal titles of that year to declare the pixel as legally or illegally mined.[26] Locations and exact shapes of Colombian legal mines were obtained from Tierra Minada.[27] The data for Peru was obtained from the Geological Mining and Metallurgical Institute of Peru.[28] 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 Minera, 2013).[29]

Finally, we add the predictions at the municipality-mineral level. We calculate the percentage of mined area of each mineral that is illegally mined. Table 3 presents the summary statistics of our predictions and a preview of our two identification strategies. According to our estimates, 84% of the gold-mining area in Colombia is exploited without a title. Although this number seems high, it is close to the 81% estimated for gold mining in 2014 using manual inspection by the UNODC (2016).

5 The effect of the reform on illegal mining

5.1 Identification strategies

We use three different identification strategies to test the predictions of our model. We use a difference-in-differences strategy comparing minerals across time within the same

[26] 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) that is closest to the vector of each pixel prediction through time.

[27] 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/

[28] Accessed through Global Forest Watch on May 22, 2016. www.globalforestwatch.org

[29] An attempt to predict the mineral mined in each area with a machine learning model was not sufficiently precise. See Figure A.2 in the Appendix.

municipality to test that illegal mining increased more for minerals with a higher tax rate. To test that mines are more likely to operate illegally after the reform, we use a difference-in-differences strategy that compares illegal mining in Colombia and Peru, before and after the reform. Finally, we use cross-sectional variation in municipality characteristics to identify the heterogeneous effect of the reform on municipalities with weak national government presence (prediction 3) and illegal armed group presence (prediction 4).

To test Prediction 1—that the reform increased illegal mining in Colombia—we use a difference-in-differences framework comparing Colombia and Peru, before and after the reform. We use Peru as the control country for several reasons: (i) It is a neighboring country that is also affected by illegal mining; (ii) Peru has levels of mineral production of the same order of magnitude as Colombia (Table A.4 in the Appendix provides more details); and (iii) Peru and Colombia simultaneously adopted an Andean Community law allowing destruction of illegal mining machinery on-site.[30] The estimating equation is:

$$\widehat{y_{mt}} = \beta_A After_t \times Col_m + \gamma_m + \gamma_t + \varepsilon_{mt}, \tag{4}$$

with the identification for β_A in the equation above coming from changes in illegal mining before and after the reform. But 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 national government's system for processing mining-title requests was down around the time of the reform. Although one might expect that the firms that

[30]Brazil was another candidate for the study, but it is not a member of the Andean Community, and thus is not affected by this law. Venezuela and Ecuador have similar levels of mineral production, but they do not have maps of legal mining titles available.

wanted to get a title 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 titles at the end of the study period, 18 months after the system was back up. That is, if a miner could not register the title while the system was down, it will not count as illegal mining in our data.[31] The second event simultaneous with the reform was a change in the law allowing destruction of illegal mining machinery on-site, rather than confiscation and a court procedure. This law was approved by both countries, and hence the difference-in-differences estimator will not be affected. A potential confounding factor could stem from a differential degree of enforcement of this law. While we do not have data on 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 are not affected by this concern. In Figure 3 and Table A.5 we test the parallel trend assumption. Colombia and Peru had similar trends and levels of illegal mining before the reform.

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 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; and also nationwide changes, like the title-request system being down. Since we are identifying the difference

[31] Mining permits were not accepted between February 2, 2011 and July 2, 2013. This was the result of 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.

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.9). We include in the Appendix other minerals as robustness. Specifically, we estimate the following equation:

$$\widehat{y_{mit}} = \beta After_t \times \alpha_i + \gamma_P Price_{it} + \gamma_{mt} + \gamma_{mi} + \varepsilon_{mit}, \tag{5}$$

where y_{mit} is the 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. $Price_{it}$ is the price of mineral i at time t. Finally $\gamma_{mt} and \gamma_{mi}$ are municipality-time and municipality-mineral fixed effects. Figure 4 suggests that the parallel trend assumption is also met for this difference-in-difference strategy.

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.

5.2 Prediction 1: Effect on illegal mining

The results of estimating equation (4) are presented in Table 4. As predicted, area mined illegally as a share of total mined area increased by 4.47 percentage points in Colombia relative to Peru, after the reform (Column 1) . The effect of the reform is larger when looking at the fraction of newly mined area that is mined illegally (Column 2 in Table 4), because this measure excludes the stock of existing mines. Another way of confirming our results is to estimate an analogous regression using titled area as the dependent variable. This measure does not depend on our mining-area predictions, and is calculated from the government's data. Results are presented in Column 3. They show a reduction

in area titled in Colombia after the reform.[32]

Finally, we study how the difference between Colombia and Peru evolves before and after the reform (Figure 3), using an event study (or dynamic difference-in-differences). As mentioned above, Colombia and Peru have similar levels (and trends) of illegal mining before the reform, but illegal mining steadily increases in Colombia (relative to Peru) after the reform.

In Tables A.6-A.7 we present different robustness checks of these across-country 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 rely on another machine learning technique to select the optimal controls. We use a Double Lasso procedure that selects controls that are relevant from a statistical point of view and are not chosen ad hoc by the researcher. 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 the price index, 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.6 in the Appendix provides the results). The procedure selects, among others, the lagged price, which makes sense given the time it takes to start up mining operations. 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

[32] As mentioned above, mining permit requests were not accepted between February 2, 2011 and July 2, 2013. This interruption may have had an impact on illegal mining. Appendix B.2 provides more details. We are unable to distinguish the effect of the reform from the effect of this interruption in mining-permit applications.

title fees and royalties. However, not all illegal mines can be made legal. For example, mining inside a national park is not allowed and hence permits within park boundaries should not be issued. Thus, we expect (and find) that the increase in illegal mining is concentrated outside national parks (Table A.7, Column 2).[33] The results using the raw probabilities that a pixel is mined, instead of a dummy, are qualitatively similar (Table A.7, Column 3).

We then estimate the same model, but using only municipalities close to the border between Colombia and Peru (Table A.7, Column 4) and find qualitatively similar results. We investigate whether the results are robust to using a different cutoff for the machine learning predictions (Table A.7, 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 coefficients with the optimal threshold. This is because the new cutoff has almost double the difference between TPR and FPR, compared to our conservative optimal threshold.[34]

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

So far, we have looked at the extensive margin of evasion, but it is possible that evasion is also present on the intensive margin through underreporting to the national government of quantity produced. As legal mines are already registered, they could evade via underreporting of production. We estimate equation (4) using reported production per

[33]Column 1 of Table A.7 replicates the results from Table 4 for convenience.

[34]Sixty percent compared to 32%. The difference between the TPR and the FPR is directly proportional to the estimate of area illegally mined. After adjusting the raw measure of the fraction of the municipality area that is mined, using the formula in Equation (8) yields similar results (Table A.7, Column 6). Appendix D.1 provides more details.

area as the dependent variable. Results are presented in Table 5. We find a positive effect for two minerals and negative for two others, although they are only significant at the 10% level for silver. This could be could be explained by the government having more difficulty monitoring the quantity extracted compared to the area mined. We argue the local authority and the miner 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.

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

The results of estimating Equation (5), the differential effects of the reform by mineral, are presented in Table 6. Column (1) confirms Prediction 2: For a mineral with a royalty tax rate one percentage point higher, there was an extra one percentage point increase in area illegally mined. Although the results are not directly comparable with those in Table 4, we can do a back-of-the envelope calculation to comparing them. The key to compare the predictions is to calculate the *average* change in the surplus of illegal mining in Colombia and compare it to the change in the surplus for each mineral (Equation 1). The value of α in Peru is 1%, while in Colombia it is around 5% on average across minerals.[35] In line with this difference of 4 percentage points across countries, the coefficient in Table 4 is around four times the coefficient in Table 6. The fact that we obtain similar quantitative results with both strategies supports the validity of both of them. However, since the within municipality-strategy requires weaker assumptions, it serves as type of robustness check for the difference-in-differences across Peru and Colombia.

We also study the evolution of illegal mining by mineral (Figure 4) using an event study

[35] In Peru, royalties are set at either a percentage of profits or 1% of sales, whichever is higher. In most cases, 1% of sales is the actual amount paid.

(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 perform robustness exercises analogous to the ones we did for the Colombia-Peru differences-in-differences. Column 2 of Table 6 presents the results with new area mined and Column 3 with the percentage of area titled. Although the results are not statistically different from zero, the results for new area mined are in line with the previous results.

Table 7 presents robustness exercises similar to those in subsection 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 — this 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 on Table A.10. Recall our preferred estimation only uses gold and coal mines, which are the two minerals mined the most according to the mining census (Table A.9). The results are robust to including the next most-mined minerals platinum and magnesium.

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 low probability of illegal mines being detected. Municipalities with low probability of detection are those with weak institutional presence of the national government. We measure it as the number of institutions (e.g., tax collection or notary office) per capita

(Acevedo & Bornacelly, 2014). As predicted, the effect of the reform is larger in municipalities with weak institutional presence of the national government (Table A.8, Column 2).[36]

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

Theoretically, we predict 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 & Bornacelly, 2014). The effect is smaller, although not statistically significant (Table A.8, Column 3). 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.[37]

5.6 Back-of-the-envelope calculations

We estimate the dollars lost through evasion 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 title fees. The coefficient of "$After \times \alpha$" is 1.46 (Column 1, Table 6). Multiplying this coefficient by the area mined and the α of each mineral, we obtain the increase in the area mined illegally due to the reform. Then

[36]Table A.8 explores heterogeneity along the following margins: institutional capacity, illegal armed group presence, and whether the municipality is a net fiscal winner or loser as a result of the reform. Illegal armed group presence is discussed under Prediction 4. Since the reform allocated some funds according to population and poverty and unemployment rates, the net impact of the reform in each municipality varies according to these characteristics. However, we do not find any heterogeneity along this margin. Since most of the royalties come from oil and gas, in general only those municipalities lose from the reform overall. However, on the margin the incentive to enforce legal mining drops.

[37]Our request for data on National Police operations to study these conjectures was denied.

we estimate the amount of mineral extracted per hectare from the 2010 Colombian Mining Census. Multiplying this by the royalty rate and price of the mineral, we obtain the dollar lost in royalty taxes. For gold it is USD 263 million and coal USD 7 million. Thus, at least USD 270 million of royalty revenue are lost with the reform. Compared to the mining royalties to be distributed—USD 594 million—this is equal to 45 cents per dollar redistributed. See Table 8 for further details.[38]

5.7 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 taxes between the mining municipality and other municipalities. Let β be the share allocated to the mining municipality, and $e(\beta)$ the function that relates evasion (e) to this share β. Given the model in Section 3 and the results, $e'(\cdot) < 0$: The greater the share allocated to the mining municipality, the smaller the amount evaded. The objective is

$$\max_{\beta} \left(\lambda\beta + (1-\beta)\right)\left(1 - e(\beta)\right), \tag{6}$$

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

[38] In addition, the total mined area is estimated to be 795,700 *ha*; therefore we estimate a 35,600 *ha* increase in the area mined illegally due to the reform. The annual title fee is equal to the daily legal minimum wage (USD 10.50) per hectare, for a total of USD 373,800 lost in title fees.

municipalities. If $\lambda < 1$, then:

$$\lambda = 1 + \frac{e'(\beta^*)}{1 - e(\beta^*) - \beta^* e'(\beta^*)} \tag{7}$$

We now substitute for the values estimated on 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, Maldonado, and Trujillo (2017) 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.

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.

Illegal mining, as a share of total mined area, differentially increased in Colombia by 4.47 percentage points after the reform. Of every dollar redistributed, around 45 cents are lost through evasion. Across minerals, we find that for every percentage point of royalty tax rate, there was an extra one percentage point increase in illegal mining after

the reform. The increase in illegal mining illustrates the difficulties of redistributing resources. Given the trend towards decentralized spending (Kim, 2018), our results point to the importance of connecting tax revenue and spending. The incentives for local authorities should be aligned with the revenue they receive.

Our results also demonstrate that monitoring illegal activity using remote sensing and machine learning is promising. Indeed, many countries have started doing so. For example, India recently announced a policy along these lines.[39] However, illegal miners could respond by resorting to underground mining, rendering monitoring more difficult.

References

Acevedo, K. M., & Bornacelly, I. D. (2014, July). *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 Minera. (2013). *Exploring opportunities: the answer is Colombia.* Retrieved from http://www.anm.gov.co/sites/default/files/Documentos/exploringopportunities.pdf

Asner, G. P., Llactayo, W., Tupayachi, R., & Luna, E. R. (2013). Elevated rates of gold mining in the amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences*, *110*(46), 18454-18459.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Banerjee, A., Mullainathan, S., & Hanna, R. (2012, April). *Corruption* (Working Paper No. 17968). National Bureau of Economic Research. Retrieved from http://www

[39] http://articles.economictimes.indiatimes.com/2016-04-12/news/72266895_1_minor-minerals-major-minerals-sand-mining

.nber.org/papers/w17968 doi: 10.3386/w17968

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29-50.

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The political resource curse. *The American Economic Review*, *103*(5), 1759–1796.

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., & Sieber, S. (2012). The political economy of deforestation in the tropics. *The Quarterly Journal of Economics*, *127*(4), 1707–1754.

Cai, H., & Treisman, D. (2004). State corroding federalism. *Journal of Public Economics*, *88*(3), 819–843.

Cullen, J. B., Turner, N., & Washington, E. L. (2018). *Political alignment, attitudes toward government and tax evasion* (Tech. Rep.). National Bureau of Economic Research.

Departamento Nacional de Planeación. (2007). *Actualización de la cartilla "Las Regalías en Colombia".*

Echeverry, J. C., Alonso, G., & García, A. (2011). *Por qué es necesaria la creaci'on de un sistema general de regal'ias.* (Nota Fiscal No. 2). Ministerio de Hacienda y Crédito Público. Retrieved from https://ideas.repec.org/p/col/000089/012223.html

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, June). Tourism and economic development: Evidence from mexico's coastline. *American Economic Review*, *109*(6), 2245-93.

Falkinger, J., et al. (1988). Tax evasion and equity: a theoretical analysis. *Public Finance= Finances publiques*, *43*(3), 388–395.

Fedesarrollo. (2014). *Mineria y medio ambiente en Colombia.*

Fisman, R., & Wei, S.-J. (2004). Tax rates and tax evasion: Evidence from "missing

imports" in china. *Journal of Political Economy*, *112*(2), 471-500.

Foster, A., Gutierrez, E., & Kumar, N. (2009). Voluntary compliance, pollution levels, and infant mortality in Mexico. *American Economic Review*, *99*(2), 191-97.

Gadenne, L. (2017). Tax me, but spend wisely? sources of public finance and government accountability. *American Economic Journal: Applied Economics*, *9*(1), 274–314.

Gadenne, L., & Singhal, M. (2014). Decentralization in developing economies. *Annual Review of Economics*, *6*(1), 581-604.

Gallego, J., Maldonado, S., & Trujillo, L. (2017). *Blessing a curse? institutional reform and resource booms in Colombia.* (https://sites.google.com/site/jagallegod/research)

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, May). Satellites, self-reports, and submersion: Exposure to floods in Bangladesh. *American Economic Review*, *105*(5), 232-36.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... 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, April). Measuring economic growth from outer space. *American Economic Review*, *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., Mejia, D., & Tribin, A. (2014). Illegal gold mining and violence in Colombia. *Peace Economics, Peace Science and 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. *Journal of Human Resources*, *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. *The Quarterly Journal of Economics*, *131*(1), 219–271.

Kim, J. (2018). Fiscal decentralisation and inclusive growth.

Lipscomb, M., & Mobarak, A. M. (2016). Decentralization and pollution spillovers: evidence from the re-drawing of county borders in Brazil. *The Review of Economic Studies*, *84*(1), 464–502.

Ministerio de Minas y Energia. (1995). *Decreto 1747 de 1995.*

Ministerio de Minas y Energia. (2003). *Glosario tecnico minero.*

Ministerio de Minas y Energia. (2011). *Resolución no. 180099 del 02 de febrero de 2011.*

Ministerio de Minas y Energia. (2016). *Política minera de Colombia.*

Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., ... Ermon, S. (2018). Infrastructure quality assessment in africa using satellite imagery and deep learning. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 616–625.

Romero, M., & Saavedra, S. (2015). *The effect of gold mining on the health of newborns.*

Sanchez de la Sierra, R. (2017). On the origin of states: Stationary bandits and taxation in Eastern Congo. *Available at SSRN 2358701*.

Slemrod, J. (2018). *Tax compliance and enforcement* (Tech. Rep.). National Bureau of Economic Research.

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, *19*(1), 25–53.

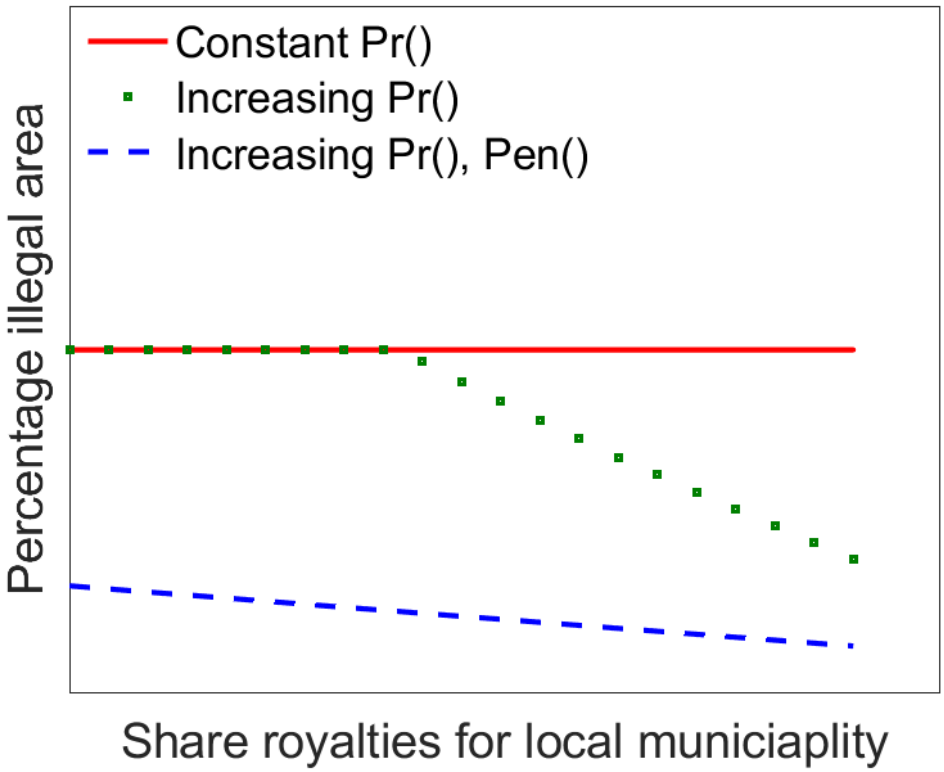
Slemrod, J., & Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In *Handbook of public economics* (Vol. 3, pp. 1423–1470). Elsevier.

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, U. N. O. o. D. C. (2016). *Colombia: Explotación de oro de aluvión. evidencias a partir de percepción remota.*

Figure 1: Share of royalties for the local municipality and illegal mining



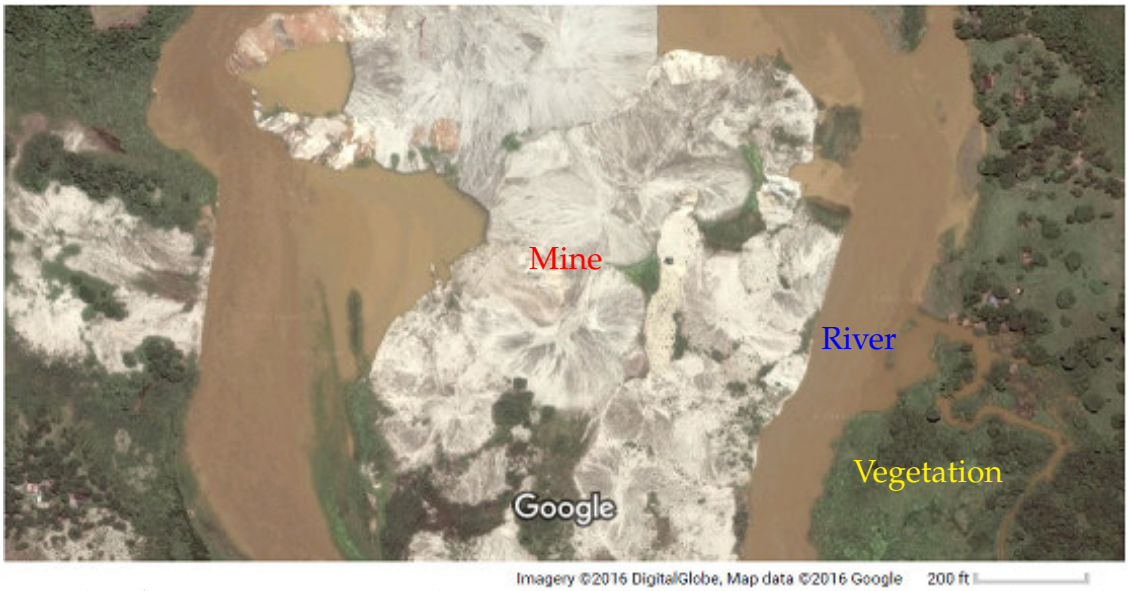
Notes: The graph shows the relationship between the share of taxes allocated for the local municipality and illegal mining, under different assumptions. The x-axis is the share of royalties allocated for the local municipality, β on the theoretical framework of Section 3. A point further to the right indicates a higher β. The y-axis is the percentage of mined area mined illegally: a higher point represents more illegal ming. The red solid line plots the case of constant probability of detection and constant penalty, 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 line represents our preferred assumptions.

Table 1: Summary statistics for municipalities used in the analysis

	(1) Mean	(2) Median	(3) Std. Dev.	(4) Min	(5) Max	(6) N
Panel A: Municipalities budget change with the reform						
% budget change	4.11	7.51	11.66	-62.5	52.3	927
% budget change if loss	-16.91	-12.52	15.02	-62.5	0	150
% budget change if won	8.17	8.24	4.13	0	52.3	777
Panel B: Characteristics Colombian Municipalities						
Population	23,471.53	12,796	36,282.56	861	377,693	927
Area (km2)	641.72	269.28	1,328.02	15.4	17,266	927
Area mining titles (km2)	0.01	0	0.06	0	1.4	927
Panel C: Characteristics Peruvian Municipalities						
Population	14,864.54	4,442	44,595.66	191	860,107	1,787
Area (km2)	627.27	207.44	1,624.63	1.9	22,184	1,806
Area mining titles (km2)	0.02	0	0.08	0	2.2	1,806
Panel D: Machine Learning Training Data						
On 2010 Mining Census	0.55	10	0.50	0	1	927
% illegal area (Census)	36.49	8.84	42.34	0	100	505
% open pit mines (Census)	78.19	1000	359	0	100	505
% illegal area | open pit (Census)	40.9	13.39	43.73	0	100	472
% illegal area (UNODC)	80.79	94.79	23.32	0	100	913
% illegal area (UNODC) | Censed	80.29	922	22.74	0.3	100	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. The rows related to the Census only have information for municipalities that were covered by the 2010 Mining Census. Population is from the 2005 Population Census. Source: CEDE panel data, 2010 Mining Census, and UNODC. In Panel C, an observation is a Peruvian municipality. There are 1,806 municipalities. Population data is from 2004. Source: National Statistics Bureau (INEI).

Figure 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). 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.

Table 3: Summary statistics, illegal mining panel

Panel A: Mineral mining Colombia

% of mined area mined illegally	Coal	Gold	Difference
Before the reform	84.25	87.96	3.71***
	(27.22)	(25.28)	(.62)
After the reform	79.99	84.01	4.01***
	(27.39)	(25.72)	(.9)
Difference	-4.25***	-3.95***	.3
	(.87)	(.67)	(1.08)

Panel B: All mining Colombia-Peru

% of mined area mined illegally	Peru	Colombia	Difference
Before the reform	87.65	87.74	.09
	(21.54)	(23.9)	(.35)
After the reform	77.22	82.18	4.96***
	(27.08)	(25.3)	(.63)
Difference	-10.43***	-5.56***	4.87***
	(.39)	(.56)	(.67)

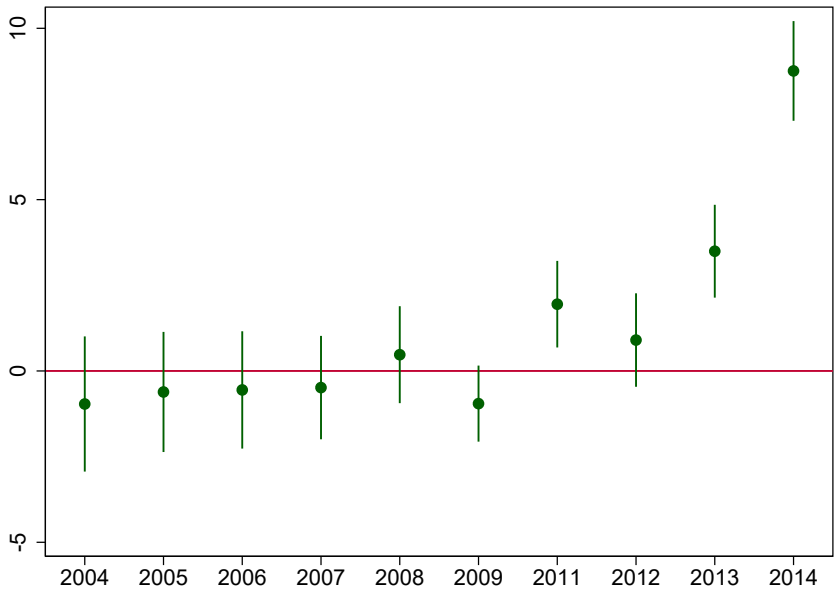
Notes: Table 3 presents summary statistics for the illegal mining panel. Panel A presents presents the mean percentage of the municipality mining area mined illegally by mineral for Colombia. In the columns the results are presented by gold/coal, and in the rows they are presented for the years before the reform (2004-2011) and after the reform (2012-2014). 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. UNODC (2016) estimates for Colombia and our own estimates of "% of mined area mined illegally" are higher than the Census estimates. Likely some illegal mines refused to answer the census or under-reported their mining area. Calculations: Authors.

Table 4: Effect of the reform on illegal mining

Dependent variable:	% mined area mined illegally		Area mining titles (ha)
	Stock	New	Stock
	(1)	(2)	(3)
After x Colombia	4.47***	5.35***	-1.22***
	(0.62)	(0.75)	(0.31)
N. of obs.	26,355	11,608	30,021
Municipalities	2,733	1,552	2,748
Mean of dep. var.	85	89	4.7
R^2	0.73	0.72	0.86

Notes: The entries in Table 4 are the estimated coefficients of Equation 4. In the first column the dependent variable is the percentage of the stock of area mined that is mined illegally in the municipality. Hence, Column 1 only includes municipalities with some 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 titles, measured in hectares. All regressions include municipality and time fixed effects. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3: Visual representation of parallel trends assumption: Colombia vs Peru



Notes: The estimates 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 excluded year. The drop for the 2012 coefficient is due to our conservative assumption of including all legal titles registered by 2014. See Figure A.5 for a version of this graph using concurrent titles.

Table 5: Effect of the reform on reported quantity

Dependent variable: Reported production by area	Coal	Gold	Silver	Platinum
	(1)	(2)	(3)	(4)
After x Colombia	-0.89	5.19	-7.88*	1.04
	(2.36)	(16.9)	(4.20)	(2.15)
N. of obs.	805	2,113	2,011	405
Municipalities	128	356	333	63
Mean of dep. var.	3.86	20.7	4.39	1.54
R^2	0.33	0.29	0.27	0.75

Notes: The entries in Table 5 are the coefficients estimates of Equation (4), where the dependent variable is the reported production of each mineral by area. 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 6: Effect of the reform on illegal mining by mineral

Dependent variable:	% mined area mineral mined illegally		% area titled
	Stock	New	Stock
	(1)	(2)	(3)
After $\times\alpha$	1.46**	1.32	0.18
	(0.72)	(1.06)	(0.11)
Price	0.21	-0.32	0.085
	(0.76)	(1.09)	(0.32)
N. of obs.	6,774	4,918	6,394
Municipalities	390	378	390
Mean of Dep. Var.	83.85	86.47	2.31
R^2	0.96	0.90	0.95

Notes: The entries in Table 6 are the coefficients estimates of Equation (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. All regressions control for the price of the mineral and 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 7: Robustness of the results to different specifications

Dependent variable:	% of mined area mined illegally					
	All	Without Nat Parks	Prob	Cutoff	Adjusted	Weights
	(1)	(2)	(3)	(4)	(5)	(6)
After $\times \alpha$	1.46**	1.29*	1.54**	0.037	-0.14	1.35*
	(0.72)	(0.72)	(0.71)	(0.21)	(0.45)	(0.72)
N. of obs.	6,774	6,760	6,866	6,746	2,014	6,772
Municipalities	390	389	395	388	156	390
Mean of Dep. Var.	83.85	83.75	83.85	91.09	80.71	83.81
R^2	0.96	0.96	0.96	0.98	0.99	0.96

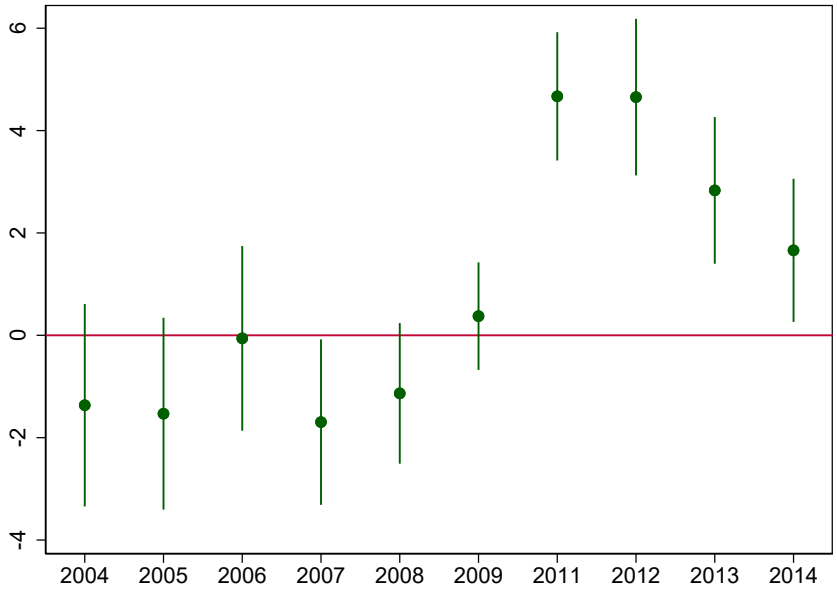
Notes: The entries in Table 7 are the estimated coefficients of Equation 5. The dependent variable is the percentage of mined area that is mined illegally in the municipality for a mineral. Column (1) is the main specification. Column (2) excludes mined areas in national parks, which cannot be legalized even if the title 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 $TPR(\rho) - FPR(\rho)$ is maximized. Column (5) adjusts the measure of area mined according to Equation (8). 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: 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	6,736.00	262.66
Coal	ton	60	864.75	5%	2,594.25	39,077.82	2,852.68	7.40

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 (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 Table 6 Column (1). Column (9) is the product of Columns (6) and (8) in millions.

Figure 4: Visual representation of parallel trends assumption: Comparing minerals across time



Notes: The estimates in Figure 4 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 indicator of gold interacted with the respective year. Point estimates and 95% confidence intervals are plotted. In short, this is the event study equivalent of estimating equation 5 and Column 2 in Table 6. The excluded year is 2010.

Appendix A Additional Figures and Tables

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[1]
Gemstones	1.5%	Variable[1]
Gold	4%	87%
Gold (alluvial)	6%	87%
Gravel	1%	67%
Iron	5%	40%
Limestone	1%	67%
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%

[1] Article 20 of Law 756 of 2002 (which modified Article 35 of Law 141 of 1994).
Notes: Tabulated data from Law 756 of 2002 (and Law 141 of 1994, which was partially modified by Law 756 of 2002).

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 % Budget Loss	0.026	-0.078	0.29
	(0.22)	(0.064)	(0.44)
Mineral price index	0.12	0.0078	0.26
	(0.12)	(0.024)	(0.31)
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

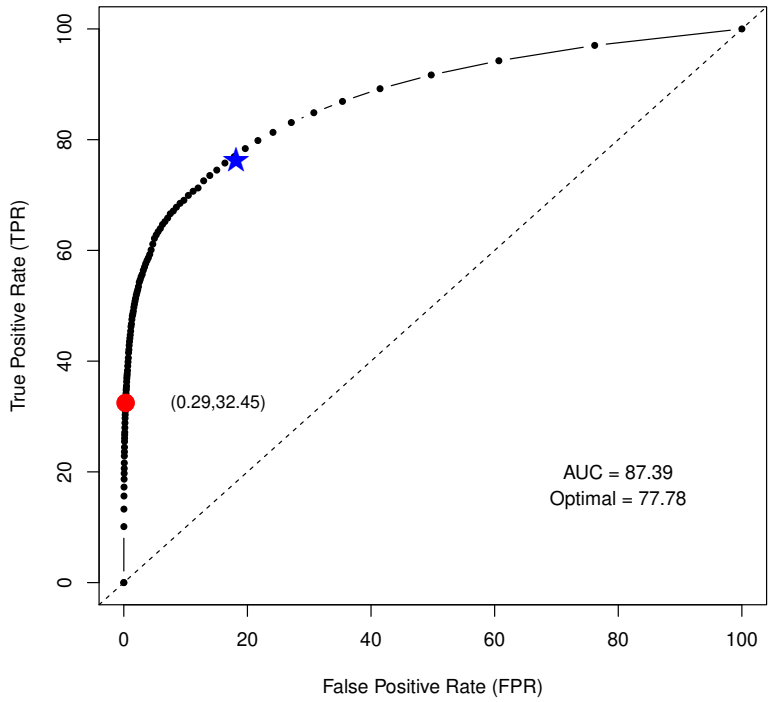
Notes: The entries in Table A.2 are the estimated coefficients of Equation 4. The dependent variable is the percentage of the stock of area mined that is mined illegally in the municipality. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Summary statistics for municipalities separated by censused status

	All	Censused	Not Censused	Difference
% Loss	-4.03	-5.14	-3.10	2.04***
	(11.6)	(10.3)	(12.5)	(0.76)
Produced precious metals	0.29	0.31	0.27	-0.042
	(0.45)	(0.46)	(0.44)	(0.030)
Oil production	0.14	0.11	0.16	0.051**
	(0.35)	(0.31)	(0.37)	(0.023)
Armed group presence before reform	0.40	0.39	0.40	0.0074
	(0.49)	(0.49)	(0.49)	(0.032)
Population	25280.0	23160.5	27072.4	3911.9
	(40628.4)	(41049.0)	(40223.3)	(2685.3)
Area (km2)	638.1	633.1	642.4	9.30
	(1330.7)	(1348.7)	(1316.7)	(88.1)

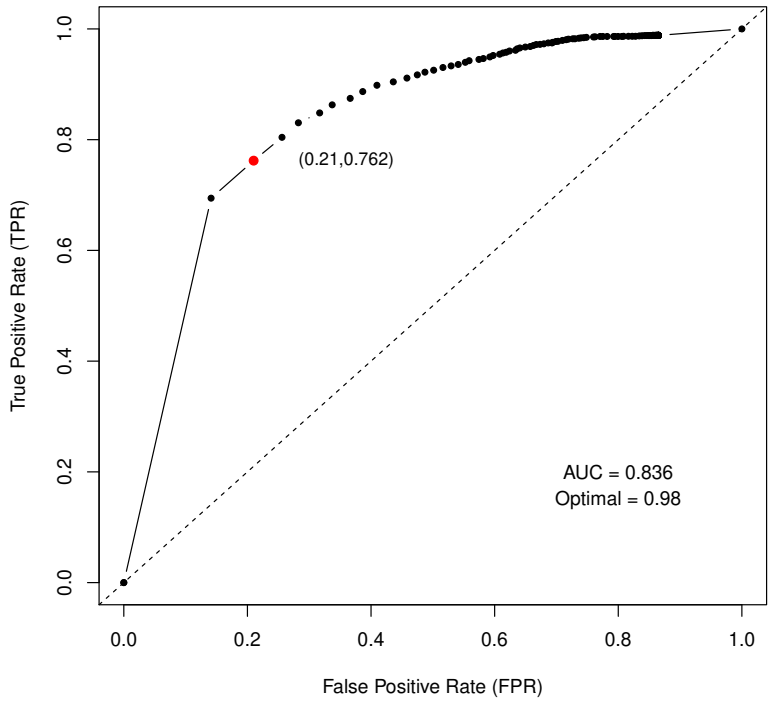
Notes: Summary statistics for municipalities, disaggregated by whether the municipality was included in the 2010 mining census. An observation is a municipality. All data comes from CEDE's municipalities panel. Calculations: Authors.

Figure A.1: ROC curve for the mining prediction model



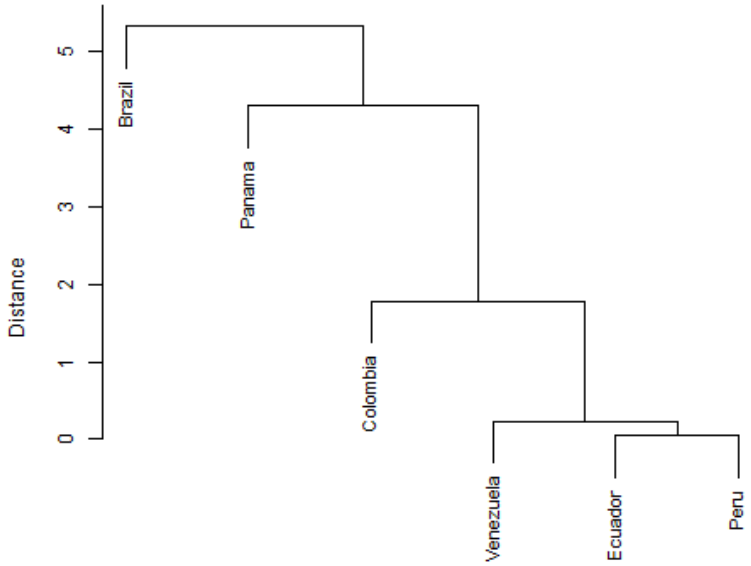
Notes: The receiver operating characteristic (ROC) curve plots the performance of a binary classification model when varying the cutoff threshold. The false positive rate (FPR)–the percentage of true no-mined pixels incorrectly classified as mined pixels–is on the x-axis. The true positive rate (TPR)–the 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.

Figure A.2: ROC curve for the mining prediction model trained to predict gold mining



Notes: See Figure A.1 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.

Figure A.3: Hierarchical clustering of neighbor countries by mineral production



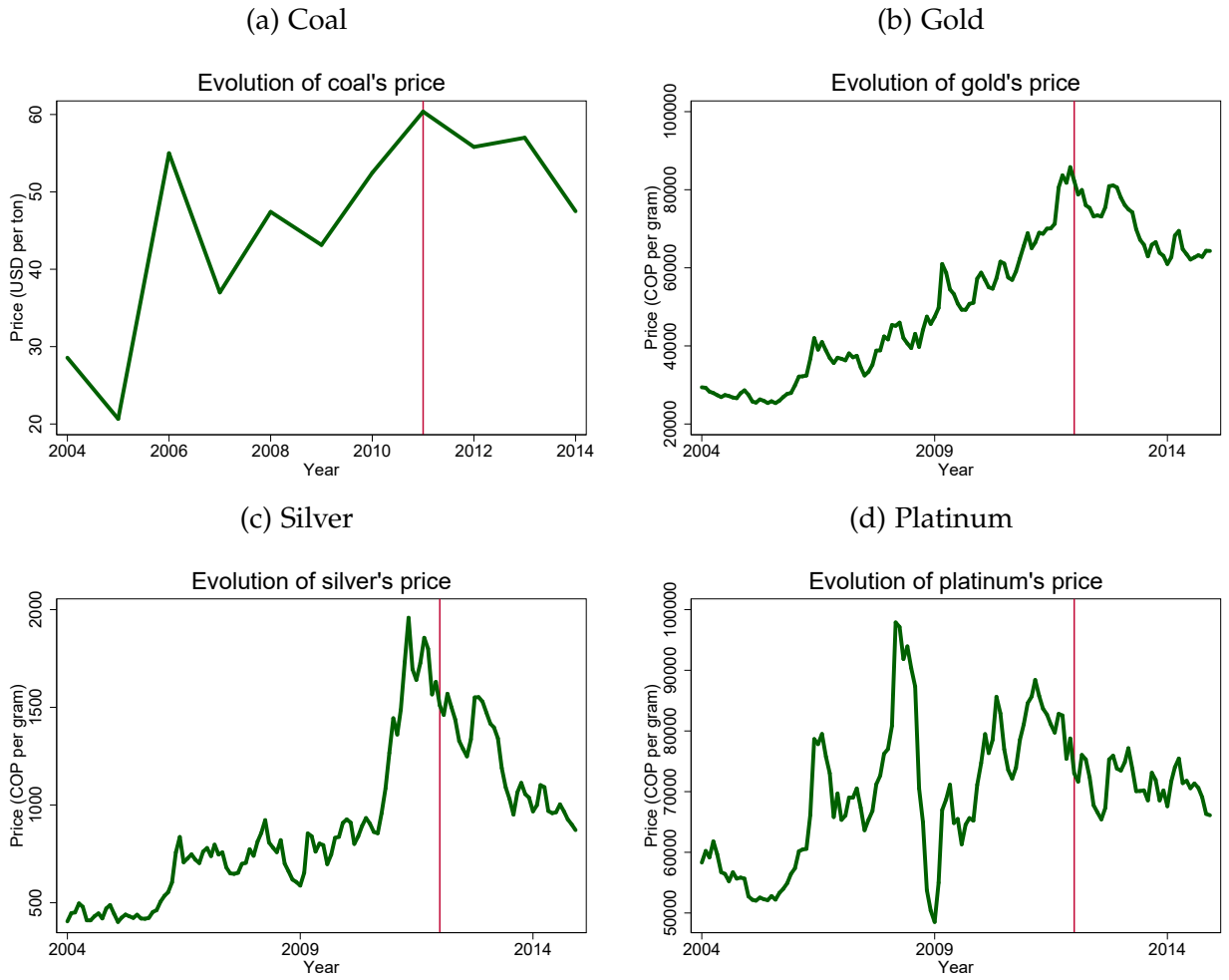
Notes: This dendrogram illustrates the distance between the normalized production vectors of each country from Table A.4. The vertical axis indicates how dissimilar the productions of two countries are. For example, Brazil's productions is relatively different from the Andean countries. Venezuela and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

Table A.4: Production of mineral commodities in 2013

Country	Alum	Copper	Gold	Iron ore	Steel	Lead	Nickel	Silver	Tin
Brazil	34,171	271	79,573	386,270	34,163	19	105	–	16,830
Colombia	–	1	55,745	710	1,297	–	70	14	–
Ecuador	–	–	2,800	–	562	–	–	1	–
Panama	–	–	2,099	–	–	–	–	–	–
Peru	–	1,286	151,486	10,126	1,069	266	–	3,407	23,688
Venezuela	2,312	–	1,691	10,583	2,250	–	6	–	–

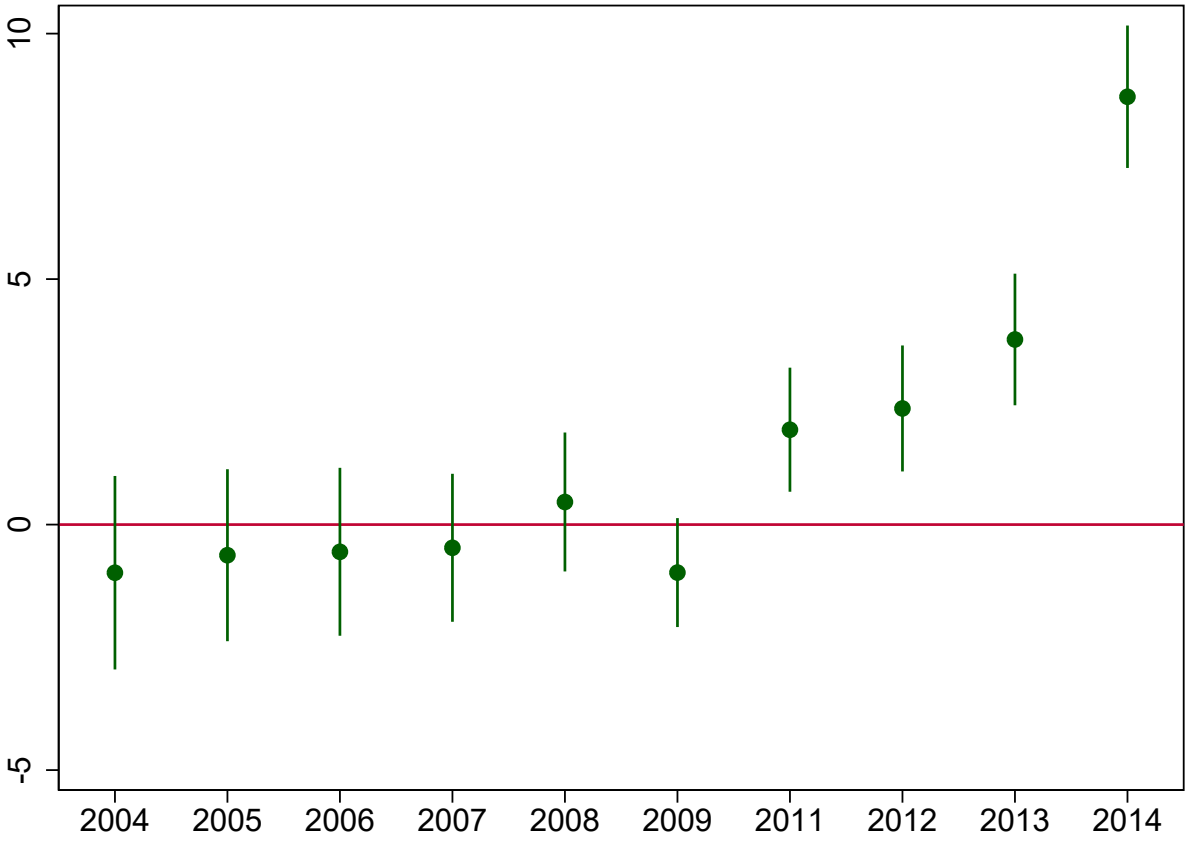
Notes: Alum stands for aluminum. Gold production in kilograms. Silver and tin production in metric tons. Other minerals in thousands of metric tons. Source: USGS http://minerals.usgs.gov/minerals/pubs/country/sa.html

Figure A.4: Evolution of mineral prices



Note: These figures show the evolution of prices for different minerals. Data for coal comes from Colombia's national government mineral information system, SIMCO. Data for gold, silver, and platinum comes from the central bank of Colombia (Banco de la Republica).

Figure A.5: Visual representation of parallel trends assumption concurrent titles



Notes: The estimates are from an event study regression for the percentage of mined area mined illegally using concurrent titles. This is analogous to Figure 3, but using the titles 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 titles granted those years or the titles of 2014, when they reopened the system to register titles. 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.5: Pre-reform illegal mining evolves similarly in Peru and Colombia

Dependent variable:	% mining area mined illegally
	(1)
Colombia × Year	0.061
	(0.15)
N. of obs.	16,222
Municipalities	2,604
R^2	0.76

Notes: The entries in Table A.5 are the estimated coefficients of Equation 4. The dependent variable is the percentage of the stock of area mined that is mined illegally in the municipality. The sample only includes pre-reform years (i.e., before 2011). 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.6: Results of selecting optimal controls with Double Lasso procedure

Dependent variable:	% of mined area mined illegally		
	(1)	(2)	(3)
After x Colombia	5.33***	4.87***	4.82***
	(0.62)	(0.64)	(0.64)
Controls			
N. of obs.	22,490	22,490	22,490
Municipalities	2,548	2,548	2,548
Mean of Dep. Var.	83.85	83.85	83.85
R^2	0.75	0.75	0.75
Controls	Main	All	DLasso

Notes: The entries in Table A.6 are the estimated coefficients of Equation 4. The dependent variable is the percentage of the stock of 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 linear trend, and interacted with 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 population. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: 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	4.47***	4.62***	5.04***	1.98***	7.54***	6.59***	5.80***
	(0.62)	(0.63)	(0.43)	(0.75)	(0.44)	(0.70)	(0.65)
N. of obs.	26,355	25,943	28,815	15,609	28,952	17,759	26,174
Municipalities	2,733	2,704	2,750	1,718	2,748	2,183	2,732
Mean of dep. var.	85.1	84.8	88.5	86.0	82.4	83.7	85.0
R^2	0.73	0.73	0.78	0.73	0.77	0.79	0.77

Notes: The entries in Table A.7 are the estimated coefficients of Equation 4. The dependent variable is the percentage of the stock of 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 title fees were paid. Column (3) uses the probability that a pixel is mined instead of the dummy of mined. Column (4) restricts to municipalities close ($<$ 1,000 km) to the border between Peru and Colombia. 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 Equation (8). 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.8: Heterogeneous effects of the reform

Dependent variable:	% mined area mined illegally					
	(1)	(2)	(3)	(4)	(5)	(6)
After x Colombia	4.40***	3.71***	4.76***	4.17***	4.67***	4.84***
	(0.62)	(0.71)	(0.73)	(0.68)	(0.60)	(1.03)
After X Weak Institutions		2.86***				
		(0.99)				
After x Armed Groups			-0.89			
			(1.01)			
After x Loser				1.33		
				(1.03)		
After x % Budget Loss					0.076***	
					(0.025)	
After x % Budget Loss if Loss						0.067*
						(0.035)
After x % Budget Win if Won						-0.098
						(0.12)
N. of obs.	26,014	26,014	26,014	26,014	26,014	26,014
Municipalities	2,696	2,696	2,696	2,696	2,696	2,696
Mean of dep. var.	85	85	85	85	85	85
R^2	0.73	0.73	0.73	0.73	0.73	0.73

Notes: The entries in the table are the estimated coefficients of Equation 4 with the added heterogeneity in each case. All regressions include municipality and year fixed effects and control for the price index. Column 1, repeats the specification of Table 4 Column 1, but restricting so that all columns have the same number of observations. 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 & 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 & Bornacelly, 2014). Since the reform allocated some funds according to population, poverty and unemployment, the net impact of the reform in each municipality varies according to these characteristics. Loser is a dummy for whether the municipality had lower transfers after the reform. Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Mineral mined according to the mining Census

Mineral	# of Mines	Percentage
Gold	3,734	27.25
Coal	2,778	20.28
Clay	2,240	16.35
Sand	1,994	14.55
Gravel	1,046	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	13,701	100.00

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

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

Dependent variable:	% mined area mineral mined illegally				
	(1)	(2)	(3)	(4)	(5)
After $\times\alpha$	1.46**	1.57**	1.16**	1.06**	0.50
	(0.72)	(0.64)	(0.54)	(0.45)	(0.38)
N. of obs.	6,774	6,774	11,905	11,905	12,680
Municipalities	390	390	551	551	556
Mean of Dep. Var.	83.85	83.85	84.26	84.26	83.98
R^2	.96	.96	.96	.96	.95
Price controls:	Yes	No	Yes	No	No
Minerals:	Gold	Gold	Gold	Gold	Gold
	Coal	Coal	Coal	Coal	Coal
			Pt	Pt	Pt
					Mg

Notes: The entries in Table A.10 are the coefficients estimates of Equation (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 controls for the price of the mineral, includes municipality-time and municipality-mineral fixed effects, and restricts the sample to gold and coal mines. The second column is akin to the first one, except that it does not include price controls. The third and fourth column, add data for platinum (Pt) mines (the third most mined mineral according to Table A.9 for which we have data). The fifth column, add data for magnesium (Mg) (the next most mined mineral for which we have data). Standard errors, clustered by municipalities, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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, Alonso, and García (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:

"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.

Based on the people's ownership right 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, in order to employ these resources (which are expected to grow in the coming years) to promote economic 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."

The constitutional reform bill (eventually known as Legislative Act 5 of 2011) was send 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 Energia, 2011). However, we believe this closure of the "mining counter" (as it is known is Spanish: "ventanilla minera") is unrelated to the immediate increase in illegal mining.

The state 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.[40] 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 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 was 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. In addition, 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; however, all minerals were subject to the closure of the mining counter.

Appendix C Theoretical Framework

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 title fees $T(Area(K))$ to the national government. But if he decides to operate illegally (I) he makes a side-payment $b(K)$ to the local authority[41] and faces a probability of the illegal mine being

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

[41] 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

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:

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

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

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 in accordance with 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.[42] 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.[43] 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

-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.

[42]The predictions on the surplus of illegal mining increasing do not require assumptions on the bargaining model. In Figure C.6 in the Appendix we are assuming Nash bargaining with constant bargaining power before and after the reform.

[43]If the bribe was paid to an agent whose payoff does not depend on the municipal budget then the reform would not have an effect on illegal mining under this framework.

associated with each type of mining, and V is the cost to the local authority if the national 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.[44] 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:

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

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

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

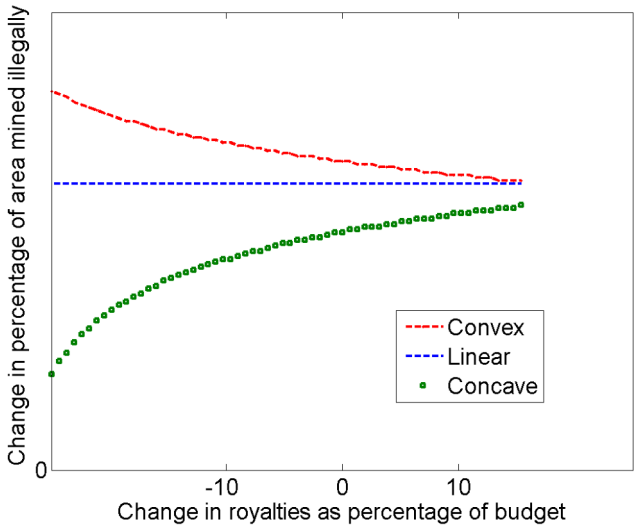
[44] In most cases the National Police destroys the machinery but does not conduct further investigation. Thus, we model V as zero.

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.[45] 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. The function $f(\cdot)$ can be convex either because local authorities capture an increasing share of the budget (Brollo, Nannicini, Perotti, & Tabellini, 2013), or because $g(\cdot)$ is convex. An illustration of this last point 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. In Colombia the median municipality spent 86 % of the revenue on "lumpy" projects such as construction of a hospital or a bridge. Figure C.6 in the appendix illustrates the predictions regarding the shape of f and the differential effect on the reform depending on the size of the budget transfer. In short, the income effect of the reform depends on the concavity of f.

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

Figure C.6: Theoretical predictions of the income effect of the reform



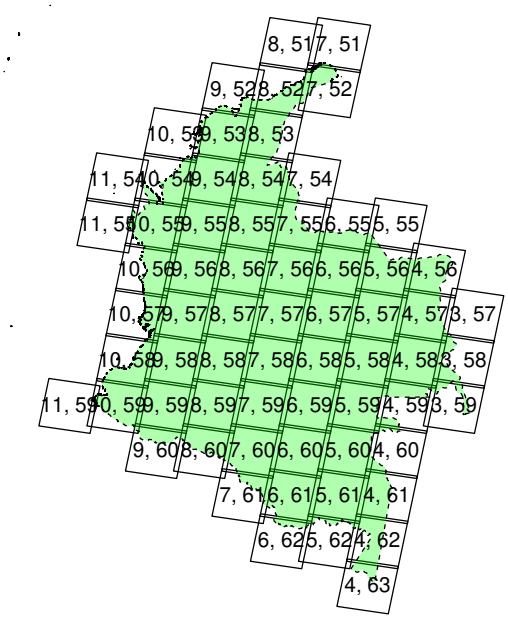
Change in percentage of area mined illegally before and after the reform, depending on the function the local authority uses to value the local municipality budget.

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.

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. Figure D.7 shows the identifiers for the images (scenes) that cover Colombia.

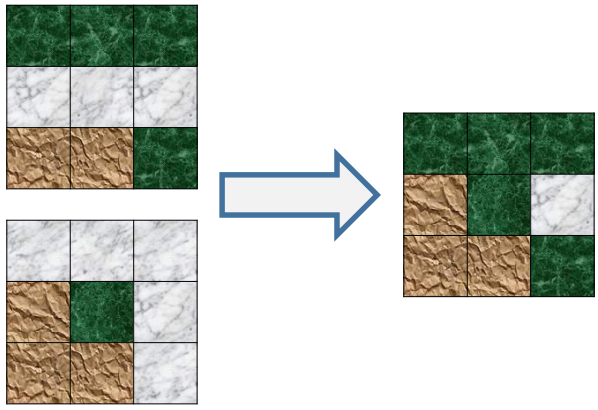
Figure D.7: Scenes (Path,row) from Landsat-7 covering Colombia



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 230MB (when compressed). A total of ∼1.5TB of raw data.

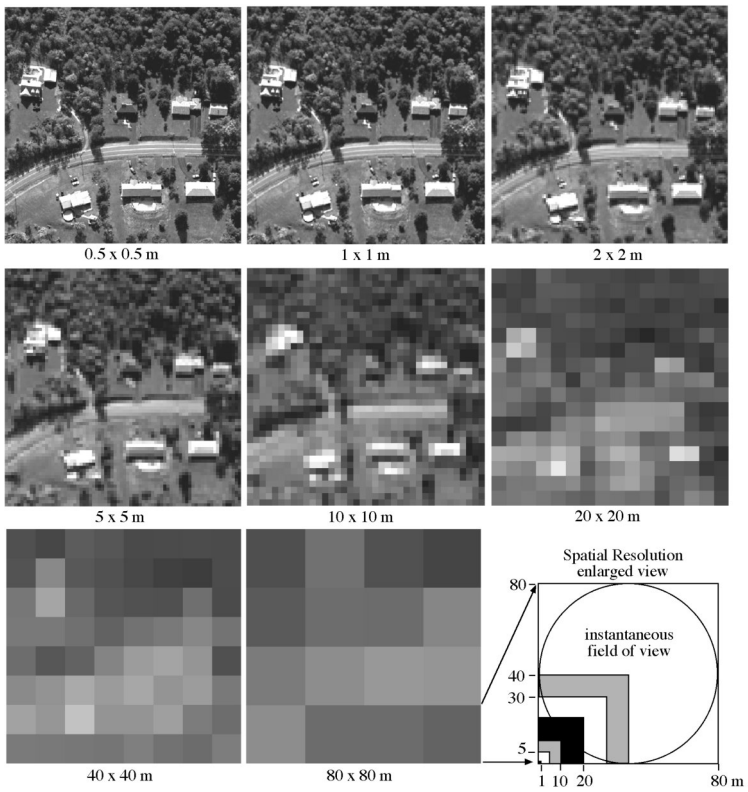
3. Given the presence of clouds, we need to construct a cloudless composite for every year. That is, we look for a cloudless image of each pixel and create a mosaic image with the cloud-free information from all images in a given year. Figure D.8 presents a toy example of how we create a cloud-free mosaic. 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.

Figure D.8: Creating a could-free mosaic



4. The resolution of Landsat is 30x30m so we cannot use shape recognition. See Figure D.9 for an illustration. Instead, we use surface reflectance information to train a machine vector algorithm.

Figure D.9



Source: (Jensen, 2007)

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.[46] We only include open pit mines, since those are the ones in which we expect to observe evidence of mining using the satellite images.

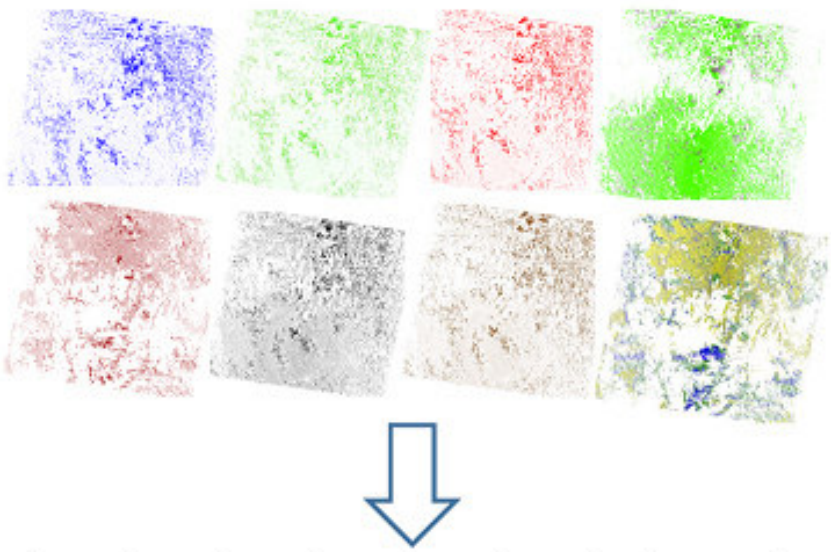
6. We confirm the presence of mines on the coordinates stated on the Census by using high-resolution images from Digital Globe (`https://www.digitalglobe.com/`). This allows us to draw the exact shape of the mine.

[46]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).

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 information[47], 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). Figure D.10 shows how the geo-spatial satellite information is transformed into a data frame.

Figure D.10: Visual representation of transforming the satellite data into a data frame



Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Loss year	Ecosystem
207	246	211.5	0	0	364	10	25
204	247	209	0	0	362.5	10	25
205.5	248.5	210.5	0	0	362	10	25
99	147	77.5	1227	495	197.5	0	25

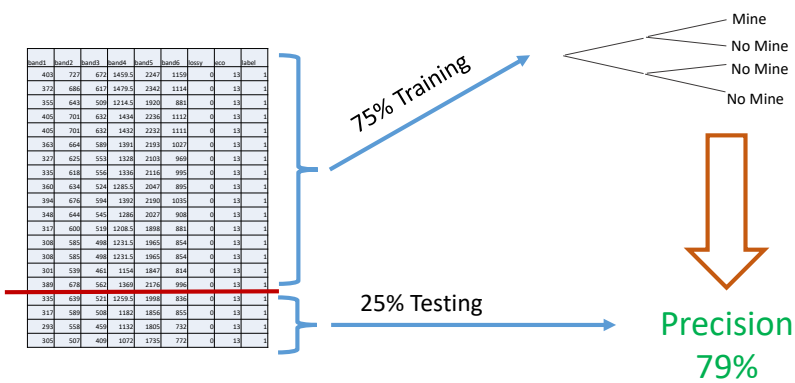
9. We exclude from the analysis forested pixels using Hansen et al. (2013)'s deforesta-

[47]Band 1 (blue), Band 2 (Green), Band 3 (Red), Band 4 (Near infrared), Band 5 (Shortwave infrared 1) and Band 7 (Shortwave infrared 2).

tion data.

10. We split the sample into training and testing sets, by dividing the country into $40km \times 40km$ 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.

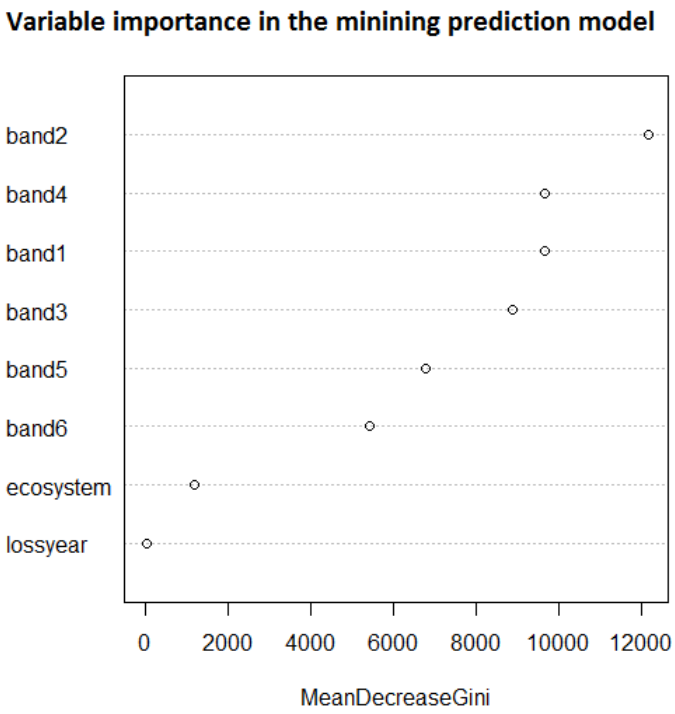
Figure D.11: Visual representation of training and testing data



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.

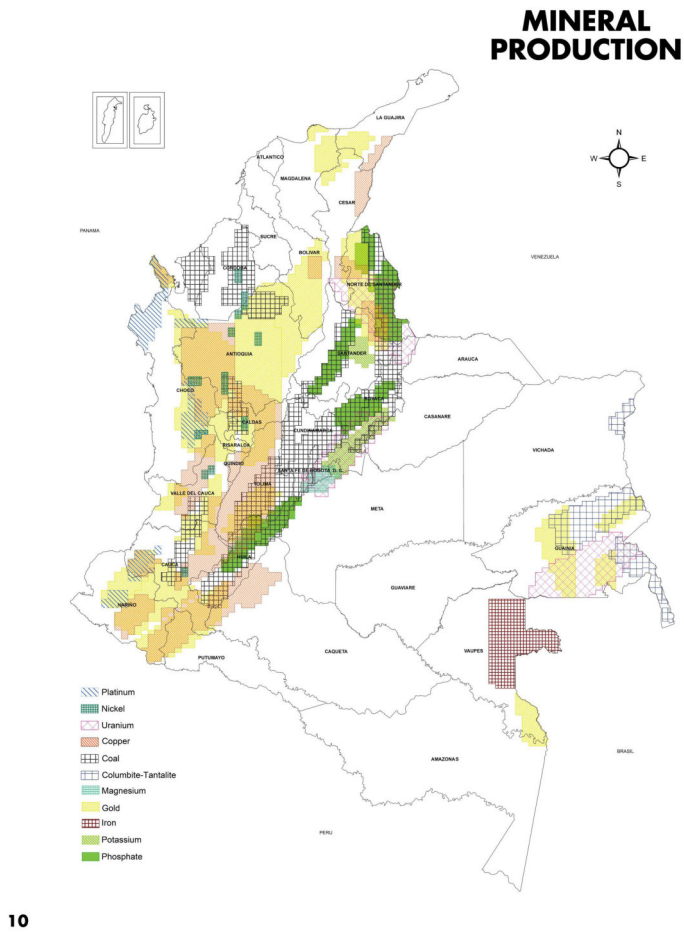
12. The random forest consists of 100 trees so it is difficult to represent its structure. However, Figure D.12 shows the relative "importance" (in what proportion of trees it appears) of each variable.

Figure D.12



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 Minera, 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 an area with a single mineral (see Figure D.13) 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.

Figure D.13: Mining potential



Source: Reproduce from Agencia Nacional Minera (2013, p.10).

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 esti-

mated measure of mining area ($\widehat{y_{mt}}$) 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:

$$\widehat{y_{mt}} = \sum_{i \in Mines} (Pred(pix_i) = 1) + \sum_{i \notin Mines} (Pred(pix_i) = 1)$$

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.[48] As the number of pixels is large, we can approximate the sum with a normal. Thus $\widehat{y_{mt}} = y_{mt}TPR + y_{Nmt}FPR + \epsilon_{mt}$, 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:

$$\frac{\widehat{y_{mt}}}{Y_m} = \frac{y_{mt}}{Y_m}(TPR - FPR) + FPR + v_{mt} \tag{8}$$

Where

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

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

[48] 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.

constant FPR will be absorbed by the municipality fixed effects.

To minimize the sum of squared errors, using formula (8), the optimal cutoff for declaring a pixel as mined is:

$$\rho^* = \arg\min_{\rho} \sum_m \left(TPR(\rho) \frac{y_{m,2010}}{Y_{m,2010}} + FPR(\rho) \left(1 - \frac{y_{m,2010}}{Y_{m,2010}} \right) - \frac{y_{m,2010}}{Y_{m,2010}} \right)^2$$

since 2010 is our training year from the mining Census. Since the fraction of 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 Figure A.1) 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 (8).

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

In this subsection we show that the independence assumption is not necessary to prove a weaker version of the law of large numbers. Let's 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's 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:

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

With some manipulation it can be shown that

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

Using Chebyshev's inequality we get the desired result.