The Private Impact of Public Maps: Landsat Satellite Imagery and Gold Exploration

Abhishek Nagaraj*

November 21, 2016

Abstract

Governments routinely invest in large-scale, scientific projects that provide basic knowledge about natural phenomena and yet the economic-value of these initiatives remains unexamined. To make progress on this topic, this study estimates the impact of Landsat, a NASA satellite-mapping program, on shaping the discovery of new deposits in the gold exploration industry. Exploiting idiosyncratic timing variation in mapping coverage, I find that information from Landsat nearly doubled the rate of significant gold discoveries, especially from junior firms and in regions with strong local institutions. The public provision of basic knowledge seems to be an important determinant of local industry performance.

*NBER and UC Berkeley-Haas (starting July, 2017). Email: `nagaraj@berkeley.edu`. I am thankful to Scott Stern, Heidi Williams, Pierre Azoulay and Catherine Tucker for their invaluable support, encouragement, and comments. Many other colleagues and seminar participants at the NBER Summer Institute, MIT Sloan Working Group and the MIT Applied Economics Seminar also provided very helpful comments. Several individuals from NASA and USGS Landsat program, and the gold exploration and remote sensing industry provided guidance. Funding from the MIT Initiative on the Digital Economy, the Kauffman Foundation Dissertation Fellowship, and the Beyster Fellowship is gratefully acknowledged. All errors remain mine.

In that Empire, the Art of Cartography attained such Perfection that ... the College of Cartographers evolved a Map of the Empire that was of the same Scale as the Empire and that coincided with it point for point.

—"On Exactitude in Science," Jorge Luis Borges

1 Introduction

Basic knowledge about the physical world has been theorized to lead to new technological innovations boosting economic growth (Romer, 1990). Governments often subsidize such basic knowledge through large-scale and expensive scientific projects such as the Human Genome Project, the Hubble Telescope or the Large Hadron Collider with the belief that the eventual, commercial benefits of these programs will boost local productivity and justify their costs (Stephan, 2012). While the association between public expenditures on basic science and growth across nations has been proposed before (David et al., 1992; Hall and Van Reenen, 2000), measurement and identification challenges have meant that there exist few causal estimates of the value of publicly-provided basic knowledge for commercial outcomes.

To shed light on this question, in this paper, I study the impact of the NASA *Landsat* satellite mapping program on shaping the discovery of new deposits in the gold exploration industry. Maps are the oldest form of publicly-provided, basic knowledge,[1] and a number of examples point to their pivotal role in shaping economic development. For example, the *Itenerario*, a compendium of publicly distributed maps (Davids, 1986) published in 1596,[2] has been associated with helping the discovery of trading relationships between the British East India company and south-asia and ending Portuguese monopoly in the region (Jefferson, 2013). Even in the 21st century, governments continue to invest significantly in mapping programs to provide basic knowledge about a region (such as its geology, weather or population) despite perennial debates about the value of these programs and consequent variation in the level of investments in basic-knowledge across countries.[3]

[1] see Borges' fantasy from the quote above

[2] containing basic knowledge about the East Indies including "very delicate nautical data that provided insight into the currents, deeps, islands and sandbanks of unprecedented accuracy for those days"

[3] See http://index.okfn.org/place/ for an overview.

In this paper, I propose that by opening the "black box" of mapping as an economic activity, it is possible to understand the role of publicly-funded knowledge on industry performance. I focus on the internal details of the Landsat program (which provided the first maps of Earth from space), in order to inform the broader question of the impact of publicly-provided basic knowledge on private-sector outcomes. While Landsat was designed to map the entire surface of the earth, in practice, there was significant variation in the timing of the mapping effort across regions. Of the 9493 "blocks" (regions of 100 sq. mile each) needed for global converage, a significant portion received satellite maps in the first few years of the program, while there was a long tail of regions which were mapped significantly later over the next decade.[4] Further, quantitative assessments and qualitative interviews indicate that significant timing differences were unintentional due to reasons like technical failures in satellite operations and cloud-cover in imagery, even though some of this variation was driven by endogenous choices on the part of program administrators who prioritized the continental USA.

I utilize this timing variation to estimate the impact of the Landsat program on commercial outcomes. While designed for its agricultural applications, knowledge from the Landsat program helped describe the basic geology of the earth's surface and aided the early-stage exploration for minerals and energy resources (Rowan et al., 1977), which is why I focus on the gold exploration sector, an industry with over $5 billion in annual exploration expenditures in 2010 alone (Schodde, 2011).[5] The main dependent variable is an indicator variable for significant new gold discoveries[6] at the block-year level obtained from a proprietary, hand-collected database of major discoveries by exploration firms between 1950 and 1990. To estimate the impact of Landsat on gold discovery, I isolate the quasi-random variation in the timing of the mapping effort using a differences-in-differences framework. Given possible concerns over selection in the timing of the mapping effort, this specification flexibly controls for differences in "prospectivity" (the "true" probability of finding resources) between different regions through time-invariant block fixed-effects, and for secular changes in the gold exploration market over time (including gold prices) through year fixed effects. A battery of tests, including testing for differences in gold discovery trends before the mapping pro-

[4] Including some blocks that were never mapped by the first generation of Landsat satellites.

[5] Conceptually, the findings from this study could generalize to exploration for other natural resources like oil and gas, copper, and uranium, even though global discovery data for these industries are harder to obtain.

[6] ranging from about 0.3-80 million ounces of gold reserves

gram was launched, region-specific time trends and excluding the United States from the analysis, help to establish the validity of the baseline specification. In addition to the main differences-in-differences specification, I also implement an instrumental-variables (IV) specification that uses the average cloudiness of different regions derived from weather data as an instrument for delays in the timing of the mapping effort to provide a robustness check for the main specification.

The implications of variation in Landsat mapping on the exploration industry are hard to predict from theory alone. On one hand, firms invest significantly in private maps motivated by the pursuit of new discoveries, and public investments in new knowledge could be duplicative or misdirected (Wright, 1983), ultimately having little impact. On the other hand, if public maps provide basic knowledge that firms find useful ex-post, but are too short-sighted or capital-constrained to invest in ex-ante (Scotchmer, 2004; Nelson, 1959), then the lack of mapping information in some regions relative to others could significantly boost regional productivity. The empirical results resolve this debate and suggest that, despite strong private incentives for mapping that have existed for centuries, the public Landsat mapping effort in the 1970s had a significant impact on the gold exploration industry. In baseline estimates, mapped regions were almost twice as likely to report a discovery when compared to unmapped regions after controlling for region and time indicators. These differences imply meaningful impacts of the mapping effort on discovery in dollar terms for affected regions—using rough estimates of discovery value (derived from data on the size of discoveries) the Landsat program led to a gain of approximately $17 million dollars for every mapped 100 sq. mile block over a fifteen year time period. For a country the size of the US (with about 3.8 million sq. miles) this translates to additional gold reserves worth about $6.4 billion USD that can be attributed to the information from the Landsat program. (See Appendix B for detailed back-of-the-envelope calculation behind these estimates.)[7]

Having found that the Landsat program had large and positive benefits in terms of overall levels of discovery, I then turn to analyzing how these gains were distributed between different kinds of market participants and regions. First, I test whether the mapping program disproportionately benefited "juniors", smaller and entrepreneurial firms in the exploration industry as compared to

[7]These estimates represent an upper bound on the value of Landsat if one assumes that public information accelerated discoveries that were bound to occur at some point in time.

"seniors," larger and more established players. I find that while juniors were making about one of every ten new discoveries before the launch of the Landsat program, in blocks that benefit from the mapping program, they report one out of every four new discoveries, a considerable increase. This translates to a 5.8 fold increase in the rate of discoveries for junior firms as compared to a factor only 1.7 for senior firms. Second, I also investigate and find significant spatial heterogenity in the impact of the Landsat program. Landsat information has a significantly larger role in increasing gold discoveries in High Income regions of the world (as classified by the World Bank), implicitly increasing gaps in productivity between the developed and the developing world. I also find complementarities between the Landsat mapping effort and the strength of local institutions (Acemoglu et al., 2000), such as strong property rights, across regions as measured by a survey of mining firms on this topic, as well as pre-1972 estimates of the strength of democratic institutions from the Polity IV project (Marshall and Jaggers, 2002). Since junior firms operate predominantly in developed regions with strong local institutions, one mechanism through which Landsat seems to have encouraged discoveries is by relaxing capital constraints for smaller firms in early-stage exploration.

The main contribution of this paper is to the economics of innovation and productivity literature on the welfare implications of public-sector investments in basic knowledge (eg: Cockburn and Henderson (1996); Lerner (2009); Bloom et al. (2002)).[8] The US government spends considerably on basic research, with R&D spending comprising between 2% and 3% of GDP (Goolsbee, 1998) since the 1960s, and has allocated almost $145 billion in R&D expenditure in 2016 (Hourian and Parkes, 2015). Despite these large investments, we know little about the causal effects of these programs on direct measures of productivity. There exists significant anecdotal evidence, such as the role of the pubclily-provided GPS system in enabling innovations in smartphone and location-based technology (Mazzucato, 2015) or the role of the funding basic biomedical research in enabling pharmaceutical innovation (Stephan, 2012). A recent stream of work has made some progress by trying to provide quasi-experimental estimates of the impact of public subsidies on private-sector patenting, an indirect, but useful measure of commercial productivity (Moretti et al., 2014; Dechezleprtre et al., 2016; Azoulay et al., 2015; Howell, 2014). For the first time in the literature,

[8] see Czarnitzki and Lopes-Bento (2013) and David et al. (2000) for an overview.

this study is able shed light on the role of public-sector investments in basic knowledge on direct measures of private-sector performance. Second, I also go beyond previous work, by highlighting the differential impacts of public-sector investments in basic knowledge on smaller and larger firms, as well as highlighting, the global and unequal benefits of US public-sector investments in basic knowledge for countries from different income categories.

More broadly, this paper also contributes to work in economic history and innovation which estimates the impact of historical investments in information and information technology on regional development and growth. The most direct connection is with David and Wright (1997), who hypothesize that American leadership in energy and minerals is driven, not simply by an exogenous endowment in these resources, but also by policy forces such as investments in geological maps and institutions like the US Geological survey. This paper is a direct test of this idea. More generally, this research is related to work that finds that information goods could be extremely beneficial for local development (Jensen, 2007; Besley and Burgess, 2001; Squicciarini and Voigtlander, 2015) and could have important implications for economic outcomes (Dittmar, 2011; Steinwender, 2014; Eisensee and Stromberg, 2007). Work on the role of institutions in influencing innovation (Furman and Stern, 2011; Moser et al., 2014) and regional development (Acemoglu et al., 2000) is also relevant.

Finally, while a large literature in economics has used satellite maps as a proxy for regional development (Henderson et al., 2012), this work contributes to a nascent literature on the economic consequences of mapping as an economic activity. Two closely related papers are Casaburi and Troiano (2016), which estimates the impact of satellite imagery on tax evasion in Italy and Williams (2013), which studies the impact the mapping of the Human Genome on innovation in the pharmaceutical industry.

The paper proceeds as follows. Section 2 helps explain some of the institutional details of the Landsat project, Section 3 describes the data and research design, Section 4 highlights key results and estimates and Section 5 concludes.

2 Empirical Setting

2.1 Landsat program

Program details: Landsat is the first and longest-running program to provide images of the Earth from space. Launched in 1972, the Landsat program has overseen seven satellite launches that have all provided "medium resolution" images of the Earth through multi-spectral cameras while revolving around the Earth at a height of about 900km above the Earth's surface. Each image from the first-generation Landsat program covers an area of about 185km $\times$ 185km, and 9493 satellite images are required to cover all of Earth's land-masses (not including Antarctica and Greenland). It is important to note that these images were significantly lower resolution than modern day satellite imagery (including modern Landsat images) which cover a much smaller land-area in a given image. For the purposes of this paper, I divide up the Earth's land surface into 9493 "blocks," each of which corresponds to a Landsat image location, and these blocks together constitute the area under study. In other words, the unit of analysis (the size of a block) is analogous to the size of Landsat imagery, rather than being artificially imposed.

The focus of this paper is the first generation of satellites in the Landsat series (Landsats 1, 2 and 3) operational between 1972 and 1983. It was not possible for NASA officials to significantly change the orbits of these satellites, however program operators usually controlled what locations were prioritized for data collection through regular instructions issued to the satellites. The Landsat satellites orbited the surface of the earth every 18 days,[9] so in principle it was possible to take repeated images of every location on earth at that frequency. However, as the Landsat literature notes "The ill-founded but frequently-held assumption that Landsat-type sensors are operated continuously as they orbit the Earth is not true" (Goward et al., 2006). In practice, as I will discuss in Section 3, many regions were left unmapped for almost a decade after the launch of the program because of difficulties of collecting, storing and relaying data back to NASA.

The Landsat imagery that was successfully collected was relayed to the Earth Resources Observation and Science (EROS) center in Sioux Falls, South Dakota which was established to collect and

[9] http://landsat.usgs.gov/about_landsat1.php

distribute these data to follow-on investigators. These data were from two sensors that captured information in both the visual spectrum as well as the electromagnetic spectrum. The EROS data center distributed these data (usually as tapes or printed images sent by physical mail) under the "open skies" mandate, which allowed governments to collect information globally, but required that the captured information be distributed at reasonable cost and without discrimination to all nations without intellectual property considerations. Given that all imagery was collected at the EROS data center, by studying the archives of this institution I am able to track information about the satellite images directly, including the location of blocks, when they were mapped, and the quality of the mapping effort including a measure of cloud-cover at the image level.[10] The prices for these data, at launch, ranged from about $10 for a 10-inch negative, to about $50 for a 40-inch color photograph (Draeger et al., 1997). According to one estimate, the cost of the program at launch was approximately $125 million (Mack (1990) pp.83).

2.2 Gold Exploration

Gold is the second most intensively explored natural resource after oil and gas, and gold mining is a complex, capital- and time-intensive process. Even though the Landsat program had implications for a number of different natural resources, my focus is on gold mining because of its relative size and importance in the mining sector, as well as for reasons of data availability.

A. Gold Exploration Technology: Organizations exploring for gold hire a team of geologists who analyze both public and proprietary mapping information to decide on a "target region." Once targets are identified, more physical, chemical, and imagery data is usually collected in the target region using both field sample collection and aerial surveys. These data are often company secrets (Hilson, 2002), obtained from archival and government mapping archives, or are collected through contractors and third-party agencies at cost. The exploration firm will use these mapping datasets to identify promising prospects, drill holes in the surface to confirm the presence of ores, and identify the economic potential of a target. Each stage of the process involves significant investments ranging from approximately $2 million per project per year for very early-stage prospecting work

[10] My primary interviews suggest that the data on the use of these images by firms was highly sensitive and has since been destroyed (Personal Communication, March 24, 2015). As such, it is unavailable for use in this research.

to figures of $5 million for advanced exploration and upwards of $1.5 billion for mine development and construction (Branch, 2009).

The payoffs for this exploration could be as much as over a billion dollars per discovery (Holdings, 2013), although there is wide variation in this number. Organizations exploring for gold include large firms that both operate mines and invest in exploration (the “Seniors”), small firms mostly funded by risk capital that are purely in the exploration business (the “Juniors”), and government geological agencies (Schodde, 2011). For the purposes of this paper, government agencies will be treated to be a part of the “Seniors” group.

B. Satellite Imagery and Exploration: After its launch in 1972, there was a gradual understanding of the utility of Landsat imagery to understand the Earth’s geology and consequently for mining. A number of geologists and academics published papers (Rowan, 1975; Vincent, 1975; Rowan et al., 1977; Ashley et al., 1979; Krohn et al., 1978) that demonstrated how satellite imagery could be used to generate targets for exploration. Landsat imagery allowed geologists to look at large swathes of the Earth’s surface that allowed them to spot large geological features that could have been otherwise invisible. Satellite maps also enabled academic and industry geologists to update maps of regions around the world to include previously unknown faults and lineaments in the Earth’s surface. Accurate knowledge of faults and lineaments is crucial for geologists because mineral resources often occur along these features. Landsat, while far from perfect, provided another important tool for firms to reduce uncertainty in the exploration process and to potentially reduce the costs of exploration. The question of whether this information was previously unknown to firms and whether it proved to be economically valuable is the empirical question that is the focus of this work.

The Landsat program needs to be understood as one among many different options for the provision of mapping information. While Landsat was the only available source of satellite mapping information, aerial information from mapping surveys conducted from airplanes were quite common (Spurr, 1954). Mapping hundreds of square miles from airplanes was considered expensive, and they were often deployed in a more precise fashion, when targets were well defined. Further, it was also possible to replicate Landsat information by launching a new, satellite-based mapping program in the private sector given the existence of commercial private-sector satellites in the telecommunications industry at this time. In fact, commercial satellite imagery did arrive in the late 1980s through the launch of the *Satellite Pour Observation de la Terre* ("SPOT") satellite system (Chevrel et al., 1981). SPOT provided satellite imagery through a commercial, "for profit" model and was launched by *Spot Image*, a French public limited company. Satellite imagery is presently provided by a number of private-sector companies, in addition to a number of separate government-run agencies. In this paper, I analyze a period in the history of this industry when the main alternative to Landsat maps was privately collected aerial images or a hypothetical privately-financed satellite mapping program.

3 Data and Research Design

Conceptually, I'm interested in four different kinds of data to help identify the relationship between new maps and the discovery of new gold deposits. All data is linked to a *block* or a 100 sq. mile patch of the surface of the earth imaged by one Landsat image. First, to quantify the timing and spatial variation in Landsat coverage, data on satellite images including mapping date, location and quality (cloud-cover) is required. Second, a comprehensive list of all major discoveries, along with discovery location and firm-type (junior or senior) is essential to quantify the main outcome variables. Third, I am also interested in covariates at the block level, including some measures of (a) the prospectivity of the block in terms of gold mining potential and (b) the local-weather conditions in terms of average cloud cover, to help assess selection issues and for instrumental variables analysis. Finally, in order to assess variation in the impact of Landsat maps by country, I am also interested in collecting measures of national income as well as the quality of different local institutional policies across different regions. This section describes the study's data collection process in further detail.

3.1 Data

A. Landsat Coverage Data: I construct data on Landsat coverage from the USGS EROS data center's sensor metadata files.[11] These data provide a list of all images collected by the Landsat sensors, including the location being imaged, the date the image was collected, and information about the quality of the image, including an assessment of cloud coverage in the image (Goward et al., 2006). I use these data to construct my main independent variables at the block-year level. First, for each block, I record the first time that it was mapped by the Landsat program to form the $Post\ Mapped_{it}$ indicator variable. Similarly, I construct a variable $Post\ Low - Cloud_{it}$ which is an indicator variable expressing whether a block has received a low-cloud image (i.e. an image with less than 30 percent cloud cover). I choose the 30 percent cutoff (Goward et al., 2006) because remote-sensing specialists indicate that images with over thirty percent cloud cover in imagery are usually unusable in practice. The results are not sensitive to the particular value of this cutoff choice (as show in Table C.2).

B. Outcomes (Dependent Variable): As far as outcome data are concerned, it is a non-trivial exercise to detect gold discoveries because of the lack of a standardized disclosure or database that tracks such discoveries. I worked with a private consulting firm to create a database that provides the date, location and additional details about economically significant gold discoveries reported since 1950. These data have been collected using press reports, disclosure documents, and other industry sources over a period of many years. While this database is unlikely to have 100% coverage, estimates suggest that about 93–99% of all valuable discoveries are included. See data appendix A for more details about this data source and for more detail on how discoveries are defined, and how date of discovery is coded. Using micro data on all available discoveries, I first match each discovery to a specific block-year using geographic coordinates in my data. Having performed this matching, I then aggregate all discoveries within a given block-year and conduct my analysis at this level. In practice, in all but 49 cases, a block-year experiences either one or zero discoveries and multiple discoveries in the same block-year are rare. Accordingly, the main outcome variable for my analysis is $Any\ Discovery_{it}$ which is an indicator variable for whether a discovery was made

[11] http://landsat.usgs.gov/metadatalist.php

in a given block-year. In total, 460 unique blocks have reported a total of about 740 significant discoveries in this period of forty years. Further, for each discovery, the database lists the names of one or more entities responsible for the discovery and a classification of whether these firms are "juniors" or "seniors"— an important dimension along which the industry classifies exploration firms. The term juniors refers to "those companies that have limited (or no) revenue streams to finance their exploration activities. Instead, the principal means of funding exploration is through equity finance." In my classification, I list "seniors" to be all other exploration companies which are not juniors. Seniors therefore include firms who finance exploration through existing revenues from production activities (usually through operating mines), and state-owned mining enterprises. Seniors can therefore be thought of as larger, older firms with pre-existing operations in the gold-mining sector.

C. Block-level Covariates: Measuring Prospectivity and Cloud Cover

In addition to the Landsat coverage data and data on discoveries, I also collect data from a number of different data-sets at the block and block-year level, to help assess selection issues and to implement my IV strategy.

First, I develop a measure of scientific interest in the area of gold geology at the block-year level. Accordingly, as a first step, I compile a list of about all 3500 publications related to gold exploration from Scopus that matched my search criteria, which provides a relatively complete index of all major scientific publications. Specifically, I search for terms related to gold mining in journals that belong to the category of "Earth and Planetary Sciences" and "Environmental Science." For each publication, using a "geo-parsing" algorithm, I identify all the geographical entities referenced in the title and abstract of the article, typically the region of the field site of the study. For example, for the article "Glacial fans in till from the Kirkland Lake fault: a method of gold exploration" (Lee, 1963), the geo-parsing algorithm would identify "Kirkland Lake" as a geological feature and a separate geocoding algorithm would provide the exact latitude and longitude of the feature, which can be used to match to a given Landsat block. Using these data and the date of the publication, I link the observation to a block-year observation in my dataset. This procedure helps me calculate the total number of gold-related publications linked to each block-year as the main covariate of interest, allowing me to create a $Pubs_{it}$ measure which captures the time-varying level of scientific research about a given block in the study period.

Second, I use the "Global Earthquake Hazard Frequency and Distribution" database (Dilley et al., 2005; CHRR and CIESIN - Columbia University, 2005), which provides a census of seismic activity to construct a block-year level measure of earthquake frequency. Geological research has shown that gold mineralization is often associated with earthquakes and related structural activity in the Earth's crust (Weatherley and Henley, 2013; Goldfarb et al., 2005). These data capture time-varying measures of seismic activity at the block-year level. I use these data combined with data on scientific publication data to create a gold "prospectivity" score (the potential of a block to contain gold) at the block-year level. In order to construct this score, I regress total gold discoveries before 1972 in a given block on total gold-mining related publications in this period as well as average number of earthquakes and then use the fitted values to predict the likelihood of gold-discovery in the post-Landsat era at the block level. As show in the scatterplots in Appendix Figure C.3, both the level of publications as well as earthquake activity prove to be reasonable predictors of gold discovery at the block level, providing confidence in the validity of the prospectivity score measure.

As a final step, I use data on average cloud cover at the block-level to create an instrument for the timing of Landsat mapping. These data are derived from the MODIS satellites by NASA and measure the average level of cloud cover at a resolution of 5km X 5km in the year 2005 (MODIS Atmosphere Science Team, 2005). These data provide a reasonable proxy for the average cloud cover that any block experiences in a given year, and thereby a good measure for the probability that Landsat map might have been obscured by clouds. I match these data and create a measure of average cloud cover percent corresponding to each Landsat block. This measure is employed in the instrumental variables specifications.

D. Measuring Income and Institutional Quality:

Finally, I also collect data describing regional income and the institutional environment in order to explore the differential impact of Landsat across these margins. First, I match each block to the respective country that it belongs to. For blocks that belong to multiple countries, I match them to the country in which most of their area lies. For each country, I collect data on the 5 "income group" classifications as defined by the World Bank including low income, lower middle and upper middle income (which I define as being in the lower tier of the income distribution) and high income (OECD and non-OECD) countries, which I define as being in the upper tier of income distribution. I examine the differential effects of the Landsat program across these two broad categories.

Second, I delve deeper into the idea that a region's economic environment is correlated with the quality of local institutions (Acemoglu et al., 2000). Notably, I am able to collect an industry-specific measure of local institutions from the "Survey of Mining Companies" conducted by the Fraser Institute.(McCahon and Fredricksen, 2014) While the survey has been conducted annually since 1997, for more comprehensive coverage, I use the 2014 edition, which contains information on over 122 different jurisdictions around the world – including provinces in major mining countries like Canada, Australia, USA, etc collected from surveys administered to over 4200 managers in the industry.[12] The survey was designed to ask managers about the level of institutional risk for exploration in different jurisdictions and regions were given a "policy perception index", based on responses to 15 different questions about the strength of different local institutions including the strength of property rights, regulatory duplication, uncertainty of environmental regulation, taxation and legal institutions etc. I rank jurisdictions by their rank on this index and test for differences in the impact of Landsat between above- and below-median regions.

A key assumption with the Fraser Institute survey measure is that local institutional conditions do not change significantly in response to variation in the Landsat effort. In order to address this issue to some extent, I use a pre-Landsat measure of institutional quality as well. This measure is derived from the Polity IV Project (Marshall and Jaggers, 2002) which codes annual information on the level of democracy at the country level. Countries are ranked on a scale of -9 to 9, where a negative number indicates an autocratic rule while positive number indicate democracy. I classify countries as having above-median levels of democracy based on whether they have a positive score in the Polity IV dataset for the latest available year before 1972.

E. Summary Statistics: Table 1 provides a list of key variables used in the quantitative analysis

[12]The survey received 485 responses (response rate of 11.5%) and firms in the survey reported a collective exploration spend of about $2.5 billion in 2014, of a total expenditure of about $4.5 billion dollars in 2014 (Carlson, 2014), and represented nearly all significant organizations in the exploration industry.

and summary statistics for the sample.

Panel A provides summary statistics for key variables that vary at the block-year level. The main outcome variable is *Any Discovery*, which is an indicator variable that is set to one if a new gold discovery is reported in a block-year. This variable is scaled by a factor of one-hundred for legibility throughout the analysis. The mean of this variable, 0.188, can be interpreted as the percentage probability that a discovery is reported in a block-year.[13] *Any Junior Disc* is set to one when *Any Discovery* is set to one and at least one discovery was reported in a block-year by a junior firm. On average, 0.038% of block-year observations report a junior-led discovery. Panel A also provides summary statistics for the key independent variables, *Post Mapped* and $Post\ Low-Cloud$ which are indicator variables that are set to one if a block has been mapped or mapped with a low-cloud image respectively by the Landsat program.

Panel B provides summary statistics for variables that do not vary over time across blocks. These data indicate that about 4.8% of the blocks ever reported a discovery, and about 3.9% of these blocks reported a discovery after 1972, the year when Landsat was launched. These data also show that the median block is mapped by a low-cloud image in 1972, however there is a long tail of blocks that remain unmapped till 1990. These data also describe the instrument, $Avg.\ Annual\ Cloud\ Cover_i$, which measures average cloud cover at the block level. The median block has a cloud cover measurement of 67.6%.

3.2 Research Design

We are interested in the impact of the Landsat maps on gold discovery. In order to identify this impact, an ideal experiment would randomly assign different quantities of Landsat imagery to different parts of the world and measure its impact on exploration outcomes. Comparing treated and control regions over an extended period of time would allow the researcher to make an assessment of the impact of Landsat data investments on gold discoveries. In this study, I use a differences-in-differences specification to approximate this ideal experiment.

[13] 99.99% of the sample reports either one or zero discoveries, and so the very small number of block-year observations that report more than one discovery in a block-year are normalized to one with this outcome variable.

A. Differences-in-Differences Specification: In order to implement the differences-in-differences specification, I first establish (in the next section) significant variations in the timing of Landsat imagery in different regions of the world. A simple comparison of the trend of gold discoveries in regions with early coverage with other regions provides a first estimate of the impact of Landsat mapping on discovery. While this comparison could be illustrative, it might be potentially misleading if regions mapped early are significantly different in terms of their potential for gold.

Motivated by this concern, the baseline, workhorse specification in this paper purges spatial differences in gold prospectivity using a block-level fixed effects approach and estimates the impact of the Landsat program on discovery using purely the variation in the timing of mapping efforts between blocks. By comparing blocks mapped early with those that were mapped late (or never mapped) I am able to estimate difference-in-difference regressions with block and calendar year fixed effects. This approach provides causal estimates of the impact of Landsat maps on discovery under the limited assumption that the timing of the Landsat mapping effort is uncorrelated with evolving understanding of the gold prospectivity of different regions. While the timing of the Landsat mapping effort was not completely random, I motivate both qualitatively and quantitatively that it was unrelated to the gold-discovery potential of different regions.

B. Instrumental Variables: While a number of specification checks and qualitative fieldwork suggest that the timing of blocks was unrelated to the evolving prospectivity of different blocks (a key assumption in the differences-in-differences estimation), I present a set of results that uses another exogenous source of variation. Specifically, I use cross-sectional variation in the average cloud cover in different regions to generate variation in the timing of cloud-free imagery being collected for a given block. The basic intuition for this idea is simple. Low-cloud regions are more likely to receive low-cloud imagery earlier as compared to regions with extensive cloud cover. This is partly because these regions are harder to image because of cloud-cover, but also because NASA administrators anticipated low-cloud cover and imaged such locations less frequently. I evaluate whether cloud cover predicts the timing of the mapping effort and consequently the timing of gold discovery. These IV estimates provide a separate way to estimate the impact of Landsat mapping on regional gold discovery and helps provide confidence in the difference-in-difference results. The empirical results section discusses different empirical strategies, empirical specifications, and results in more detail.

C. Additional Robustness: Finally, I also provide a number of additional robustness checks for the baseline analysis. This includes excluding the USA from the analysis (because NASA was focused on good coverage in the US), excluding USA, Canada and Australia (the top three producers of gold) as well as excluding blocks that had already reported discoveries before 1972 from the analysis. Further, I also conduct a placebo exercise where I replace the year in which the block first received a Landsat image with a randomly generated year after 1972 as a further robustness check. I also test whether the results are driven by the choice of the start and end year for the panel (1950 and 1990), by the definition of "low-cloud" in terms of cloud cover percentage, by blocks mapped in the first year of the program as well as by adjusting for spatial dependence in the clustering of standard errors. Finally, I also implement a cross-sectional regression specification that analyses the relationship between Landsat delays and gold discovery without block fixed effects, but with detailed controls for gold prospectivity and reasonably stringent subregion, continent or block-group fixed effects. The empirical results section discusses these robustness tests in more detail.

3.3 Landsat coverage and selection issues

Before the validity of the differences-in-differences specification is established, it is important to investigate the assumption that the timing of the mapping effort was unrelated to the changing prospectivity of different regions in terms of their gold potential. While the block and year fixed effects control for static, time-invariant factors that affect discovery, the possibility that mapping was correlated with changing trends in gold potential remains a significant concern. In this section, I establish that the concern that the timing of the mapping effort was related to gold discovery trends is unlikely to be a major impediment in my setting using both interview and archival data, as well as quantitative selection analysis.

A. Qualitative Evidence:

A few recent studies analyzing Landsat holdings (Draeger et al., 1997), have found significant gaps in coverage and have investigated the reasons for these gaps. The overarching conclusion from these studies is that the gaps are likely related to (a) administrative decisions to focus on complete coverage of the continental United States and (b) technical failures in mission operations (Goward et al., 2006). As this paper notes, this variation was both unexpected and unnoticed till quite recently.

> What we had not expected to see in the coverage maps were the variations in the geographic coverage achieved from year to year ... As we investigated further, we found that technical issues such as the on-board tape recorders on Landsats 1, 2, and 3, which typically failed early in the missions, may have caused the annual or seasonal gaps in coverage .. the options for down-linking acquired data to the ground stations decreased as on-board Tracking and Data Relay Satellite (TDRS) Ku-band and direct-downlink X-band systems started failing. (Interview, 8th April 2015)

In addition to these scientific studies and reports, I also interviewed some of the key program administrators who were responsible for Landsat mission planning in the 1970s as part of this study. They confirmed that, while it was possible to program the regions where Landsat would collect data, significant variation in Landsat coverage was due to technical errors:

> All the satellites relied on recorders, wideband videotape recorders, they were all cassette tape. If you remember cassette tape, they would get worn-out, they often failed before their intended design life ... we have a lot of data that is listed as not quality. (Interview, 6th February 2015)

They also indicated that the Landsat planning team was deliberately insulated from firms in the private sector (like exploration companies) because, as a government agency, NASA did not want to be seen to be catering to the needs of a select few. They stated that the mission was primarily focused on complete coverage of the United States, and while global coverage was desirable, the program administrators acknowledged "that's the one that ended up suffering the most" (Interview, 8th April 2015).

Finally, in addition to the specifics of mission planning (which were unrelated to gold exploration) and technical failures, variation in coverage was also due to poor quality of satellite images that were rendered unusable due to significant cloud cover. To this day, a central challenge in using satellite imagery is the presence of clouds between the satellite sensor and the land surface being imaged. According to geologists and remote-sensing scientists, an image must have less than 30 percent cloud cover (Goward et al., 2006) to be seen as useful for analysis. This requirement meant that regions that are cloudier than usual were often harder to map than regions where cloud cover is not an issue. For example, one of my interviews validates that some regions were either not mapped, or were mapped at a later point in time because it was difficult to get cloud-free imagery:

> our ability to predict clouds [is limited] ... everything comes in big fronts, especially around the equator, where there are convector, pop-up storms, and no predicting when or where they are, after a few tries you might end up with only about one or two scenes that are very clear. (Interview 22nd November, 2014)

These facts suggest that the timing of the arrival of cloud-free maps seems to follow an even more random process than the timing of the mapping effort. Motivated by this fact, in the differences-in-differences research design, I will use the timing of the arrival of cloud-free imagery (in addition to the timing of the mapping effort) to disentangle the role of Landsat from other confounding factors. Further, the IV specification will also use the average cloudiness in a given region to instrument for this timing variable.

B. Quantitative Evidence:

While the interviews and the archival analysis are helpful in establishing that the timing variation in mapping activity and the arrival of cloud-free imagery was not directly linked to trends in the gold exploration industry, in this section I test these claims quantitatively.

A simple time-series comparison of average gold discoveries between blocks that received Landsat coverage earlier as compared to other blocks is represented in Figure 1. Blocks "Mapped Early" are all blocks that received a Landsat image for the first time before 1974. As the figure illustrates, discoveries in blocks mapped early and late had fairly flat and parallel growth rate before 1973 when Landsat data was made available. After this date, both time series appear to show an increase in gold discoveries, but the blocks mapped early show a quicker rate of growth as compared to blocks mapped late. This analysis provides some preliminary evidence to suggest that early Landsat mapping had a large impact on gold discovery. However, while there do not seem to be any trend differences in trends between early and late blocks before the launch of the Landsat effort, there remain some differences in levels that could be a concern for the analysis.

Table 2 Panel A, further investigates these level differences between mapped-early and mapped-late blocks. The data indicate that mapped-early blocks are slightly more likely to report discoveries before 1972 (even though this difference is not statistically significant), have higher prospectivity scores, and are more likely to have gold-mining related publications and greater earthquake hazard. While in theory these level differences are not directly problematic for the difference-in-difference specification, the baseline DD specification will include fixed effects at the block-level to flexibly control for all cross-sectional time-invariant differences between early and late blocks.

To test the validity of the IV specification, Table 2 Panel B, compares regions that typically have a low level of cloud cover, with regions that are usually cloudy. Contrary to Panel A, these data show that these two types of blocks are comparable in the cross-section in terms of the number of discoveries before 1972, and the prospectivity score.[14] This analysis provides some preliminary evidence to suggest the validity of the IV strategy. The results section will investigate the exclusion restriction more formally.

Combined, the qualitative and quantitative data provide confidence in the validity of both the difference-in-difference specifications that exploit the differential timing of the mapping effort and the timing of cloud-free mapping as well as the instrumental-variables strategy that exploits the cloudiness of different regions.

[14] There seems to be some difference in the number of publications. However, the difference is in the opposite direction to what would be a concern for the IV specification, i.e. cloudier regions have a higher level of publications as compared to less cloudy regions.

4 Results

4.1 Did Landsat Boost Discovery?

A. Baseline Regression Specification: I now analyze the impact of Landsat coverage on gold discovery in a regression framework. The sample is constructed as follows. I divide all of the landmasses on Earth into 9493 blocks, each of which corresponds to a Landsat imaging location. For each block, I collect data on gold discoveries between 1950 and 1990. I then construct measures of Landsat coverage as illustrated in the previous section.

I use OLS to estimate the following regression specification using the block-year level panel:

$$Y_{it} = \alpha + \beta_1 \times Post_{it} + \gamma_i + \delta_t + \epsilon_{it}$$

where γ_i and δ_t represent block and time fixed effects respectively for block i and year t. $Post_{it}$ equals one for all blocks after they have either been mapped or have received an image with low-cloud cover.

This specification compares the difference between blocks that have received mapping information, with blocks that have yet to receive maps, in a differences-in-differences framework. If blocks that receive early coverage following the Landsat launch do indeed report more gold discoveries earlier, then we should find that the difference-in-difference estimate β_1 is positive. This specification also includes controls for block and year level fixed effects. Block-level fixed effects difference out level differences in underlying potential for each block (a significant concern in this setting) and year-level fixed effects difference-out time-varying environmental factors, such as gold price, which could significantly influence discovery. It should be noted that there was a large run up in gold prices in the early 1970s[15] which the year fixed effects help to control for, among other time-varying global trends such as improving extraction and exploration technology. Further, the distribution of the outcome variable (unreported) suggests that most block-years report either no discoveries or at most one discovery. Accordingly, the main outcome variable is operationalized as an indicator zero/one variable, $Any\ Discovery_{it}$, and I estimate all regressions using linear ordinary-least-squares (OLS)

[15] http://www.macrotrends.net/1333/historical-gold-prices-100-year-chart

models. All my specifications cluster standard errors at the block level, given the concern that discoveries within blocks are likely to be correlated over time. In additional robustness checks, I include more general clustering that takes seriously spatial proximity between different blocks and find that the results are generally robust to these additional restrictions (see Table C.5).

Table 3 presents estimates from this regression for both the $Post\ Mapped_{it}$ and $Post\ Low-Cloud_{it}$ variables. Columns (1) and (2) do not include block fixed effects, while columns (3) and (4) include them. The coefficients generally reduce in size after controlling for block fixed effects, indicating their importance in this setting. The results indicate that after controlling for block and year level fixed effects, there seems to be a positive impact of Landsat coverage on gold discovery. Specifically, the estimate of β_1 indicates an increase of between 0.152 - 0.164 percentage points on average of making a gold discovery after the Landsat mapping effort, a significant increase given that the baseline rate of discovery is about 0.19%. This represents almost a doubling of the rate of discovery in treated regions. As an additional robustness check, I also estimate the above specification using negative binomial models, given the skewed distribution of the outcome variable. Results from this analysis are presented in Appendix Table C.7. These results also echo the results from the OLS models. The coefficient of interest in the model using the $Post\ Low-CLoud_{it}$ variable and including both block and year fixed effects, for example, is about 0.609, which translates to about a 83% increase in the probability of discovery, similar to estimates from the OLS models. The baseline results and the robustness check therefore confirm the main hypothesis that the Landsat mapping effort had a significant impact on industry performance. In order words, despite strong private incentives for gold discovery and the significant investment made by firms in private knowledge before the 1970s, the Landsat program boosted discoveries in regions that benefited from publicly-provided mapping information.

B. Time-varying Estimates: I then turn to estimating the time varying impact of Landsat coverage on gold discovery. Specifically, I estimate

$$Y_{it} = \alpha + \Sigma_z\ \beta_t \times 1(z) + \gamma_i + \delta_t + \epsilon_{it}$$

where γ_i and δ_t represent block and time fixed effects respectively for block i and year t, and z represents the "lag," or the years relative to a "zero year," which marks the year when a block was first mapped with a low-cloud image.[16]

Figure 4 presents estimates of β_t from this regression, which measure the difference between treated and control blocks for every lag year. The dotted lines represent 95-percent confidence intervals. The figure is illustrative for three reasons. First, there seem to be no pre-existing differences in trends between the two groups, suggesting that discoveries in treated blocks were evolving at a similar level as compared to control blocks. Second, there seems to be a large and persistent increase in the number of discoveries in the two groups, confirming the effect detected in the baseline estimates. Finally, this increase seems to appear after a lag of about seven years. This delay accords well with my interviews with gold exploration companies who certify that Landsat represents early-stage exploration and is typically followed by many years of further exploration, and also with reported accounts of discovery timelines in the gold exploration industry (Branch, 2009).

C. Instrumental Variables evidence: Now, I turn to analyzing the impact of Landsat coverage on discoveries using cloud cover at the block level as an instrument for Landsat timing effort. The primary concern with the differences-in-differences estimate is the assumption that the timing of the Landsat mapping effort was unrelated to the changing prospectivity of different regions. The time-varying analysis presented in Part B helps alleviate this concern significantly. As a further robustness check, this section uses instrumental-variables estimation to understand further the role of Landsat mapping on discovery.

Consider Panel A in Figure 5. This figure plots the average cloud cover at the block level collected from weather databases and the average year in which a block was first mapped with a low-cloud image using binned scatterplots. Blocks are binned according to average cloud cover measured upto two decimal digits (i.e. 0.01 to 0.99) on the x-axis, and average first low-cloud year for each of these approximately hundred groups are on the y-axis. Panel A shows a strong positive correlation, indicating that regions with higher levels of cloud cover are more likely to receive mapping information later rather than sooner. Similarly, Panel A, Figure 2 shows the relationship between cloud cover and the average value of the Post Low-Cloud indicator variable. This scatterplot also

[16] For the small percentage of blocks that never get an image, z is consistently set to zero.

confirms the intuition that regions with more cloud cover on average have poorer image quality as compared to less cloudy areas.

These data suggest that cloud cover might be a potentially good instrument to understand the role of the Landsat mapping effort on gold discovery. However, for cloud cover to be a valid instrument, it needs to satisfy the exclusion restriction. In other words, cloud cover must predict gold discovery only through its role in influencing the quality and timing of Landsat mapping, rather than through other channels. For example, if increasingly earthquake-prone regions are also more likely to be cloud-free, then we might doubt the validity of the exclusion restriction because geological research suggests that earthquake-prone regions are also useful targets for gold exploration. Figure 5 Panel B tests whether the exclusion restriction seems plausible, although it is hard to test it formally. Panel B, Figure 1 analyzes the relationship of the prospectivity score of a block calculated based on the number of publications and the earthquake hazard index with the cloud cover of a region. Panel B, Figure 1 shows that there is a slightly positive or flat relationship between predicted prospectivity and cloud-cover. Similarly, cloud cover does not predict the number of discoveries of gold pre-1972, as indicated by the scatter plot in Panel B, Figure 2, a more direct test of the exclusion restriction. These plots provide confidence that the IV specification is likely to satisfy the exclusion restriction.

Table 4 provides estimates from a differences-in-differences specification similar to the baseline, where the $Post_{it}$ variable is instrumented by the average cloud cover at the block level. Note that while the cloud-cover variable is time-invariant and would be absorbed by the block-leve fixed effects, I interact this variable with the $Post_{it}$ indicator variable, and use this interaction as the IV in my estimation. Column (1) suggests a strong first-stage between the two variables, i.e. a higher value of cloud cover indicates that the block is likely to receive a low-cloud image later rather than sooner. The IV estimates are presented in Column (2). This estimate is about 1.126— much greater than the baseline estimate. This estimate implies that compared to the average rate of discoveries in a block-year, mapped blocks are about 6 times more likely to report a new discovery, a large and economically significant effect. This large difference between OLS and IV estimates could be attributed to differences in the local average treatment effect of the IV specification. More specifically, it is is possible that the variation arising from cloud-cover is more localized than the satellite mapping variation. This localized variation combined with large and postive inter-block spillovers could generate the patterns I observe, although I'm unable to test this hypothesis directly. In sum, I interpret these results as providing a validity check for the baseline specifications, but continue to emphasize the baseline estimates because they offer more conservative estimates of the impact of the Landsat mapping on gold discovery.

4.2 Additional Robustness Checks

A. Excluding Certain Regions: Appendix Figure C.2 plots the variation in the dependent variable exploited in the analysis. Panel A plots the spatial variation in the timing, and Panel B plots the time-series variation. As is evident from this figure, the spatial variation in early mapping is not completely random. The United States is unsurprisingly mapped early in the program, confirming the qualitative findings. Further there seems to be some evidence that parts of the USSR, were mapped early, while others were left off of the map till considerably later. Given the strategic interactions between the US and the USSR during the time period of the study, it is possible that this variation could also be problematic. In other parts of the world, notably Asia, Africa and South America, the variation seems to be more idiosyncratic. In order to address concerns that regions with problematic variation are driving the effect, I repeat the analysis for different subsamples of the data excluding these regions in turn, as described below.

First, one might be worried that the results might be driven by the US – which is both a leading producer of gold and was well covered by the Landsat program. In Table C.1, Panel A, I repeat the baseline specification completely excluding the US from the analysis. The size of the coefficient reduces slightly, but remains positive and significant. Similarly, one might be worried that the leading three producers of gold, Australia, Canada and the US might be driving the results. Panel B presents estimates without these three countries, and while the co-efficient size reduces considerably, it is still large in magnitude and statistically significant. Similarly, given the variation over the erstwhile USSR could be considered problematic, in Panel C, I present estimates excluding blocks in the USSR altogether. In a similar spirit to Panel B, Panel D presents estimates that exclude blocks where discoveries had already been reported before 1972. The idea is that if these blocks were imaged early (presumably due to selection issues), then we might be worried that the Landsat effect is capturing selection. However, in each of these four subsamples, excluding these blocks leaves the estimates virtually unchanged, indicating that most of the overall effect is driven by new discoveries in previously unexplored blocks in a variety of different regions around the world. Overall, the estimates in Panels A, B, C and D help to establish that certain possibly problematic regions are not dramatically influencing the estimates.

B. Excluding Blocks At Program Launch: Apart from the cross-sectional variation, Figure C.2, Panel B also highlights the time series variation being exploited in this paper. As is evident from this graph, approximately 75% of the globe is mapped in the first two years of the Landsat program after which there is a significant delay for blocks which were not mapped in this early period. This is likely to due technical failures forcing Landsat administrators to severely restrict data collection and failures in data transmission after the first two years of launch (Goward et al., 2006). While this pattern is not directly concerning for the analysis, it would helpful to establish the robustness of the finding to excluding the blocks mapped early in the program, thereby purely exploiting the variation in the long-tail of the mapping effort in the later years of the Landsat program. Accordingly, Table C.4 estimates the baseline specification excluding all blocks mapped in the first year, and the first two years of the Landsat program. When 1972 is excluded the estimates decrease slightly in size and remain significant. When the first two years of the program are excluded the estimates decrease to about 0.09 (from about 0.15), a significant reduction, and the standard errors increase as well. However, despite the large reduction in sample size, the coefficients remain positive and economically and statistically significant. This is reassuring because the impact of the Landsat program is preserved even when focusing on the variation in the timing of the mapping effort after the first two years of program launch, after almost three quarters of the dataset has been dropped.

C. Placebo Test (Tree-Cover and Randomized Independent-variable):In addition to excluding certain countries and blocks, the subsample analysis in Table C.1 also allows for a placebo test of the Landsat program. In particular, while Landsat is useful to divine the geology of a region for gold exploration, its utility is severely diminished in regions where tree cover obscures details of the land surface underneath, a fact that was also validated by my interviews with remote sensing experts. Accordingly, I use a dataset of global tree-cover to extract blocks which contain significant tree-cover and estimate the impact of Landsat on this limited group only. The results from this analysis are presented in Table C.1 Panel D. As the estimates show, when restricted to regions with significant tree-cover, the impact of the Landsat program is close to zero and insignificant, implying that the informational content of Landsat maps was an important channel through which new discoveries were enabled. In a similar spirit, Table C.6 presents estimates from a placebo specification where for each block I generate a randomly generated year between 1972 and 1990 in which I assume it has been mapped. Using this alternate Post Low-Cloud variable, I repeat the analysis and estimate the effect of Landsat to be close to zero and insignificant. This is reassuring because the main estimates do not seem to depend on the particulars of the specification, rather representing plausibly causal impacts of the mapping information.

D. Cross-Sectional Specification:

A key benefit of the panel specification used in the analysis is the ability to include flexible block and year fixed effects. However, it is also possible to implement a cross-sectional specification to estimate the impact of the Landsat program on gold discovery. This specification under-emphasizes small timing differences between blocks in terms of mapping, and instead simply relates overall delays to the probability that any discovery is reported in the 20 year period following the launch of the Landsat program. The baseline specification is of the form $Y_i = \alpha + \beta_1 \times Delay_i + \gamma_i + \epsilon_i$, where Y_i the main outcome variable is an indicator for whether any discovery was made in a given block i after the launch of the Landsat program in 1972 till 1990, $Delay_i$ is the total difference between the year in which a block was mapped with a low cloud image and 1972, γ_i represents spatial fixed effects. Here γ_i no longer represents block fixed effects, but rather larger spatial categories (such as the continent, "subregion" or block-groups). This specification estimates the impact of delays in mapping on gold discovery within these spatial units. The estimates are presented in Table C.9 and the footnotes include more details on the estimation. The results support the basic conclusion, that delays are associated with lower probability of discovery for Landsat blocks, even in the cross-sectional specification.

E. Region-Specific Time-Trends:

An alternative story that could cast doubt on the finding could be that mapping coincides with institutional improvement in various regions (e.g. the former Soviet Union), and confounds the direct effect on gold exploration. While this explanation seems implausible because it would require local conditions to change precisely around the timing of the Landsat mapping for a large number of blocks, I am able to directly test it using region-specific time-trends in the regressions specification. Instead of including a common year-specific indicator variable at the global level, this specification includes separate year-specific dummies for different sets of regions around the world.

Estimates from such a specification are presented in Appendix Table C.8 and allows us to evaluate the robustness of the baseline results to region-specific time trends. Specifically, I estimate three separate models that split the globe into regions in different ways. First, one might be concerned that countries in different income categories around the world are modernizing in different ways, possibly influencing gold exploration. Accordingly, I identify blocks as belonging to one of five different income categories (High income: OECD, High income: non-OECD, Upper middle income, Lower middle income and Low income) and introduce separate annual time trends for each of these regions. Second, I perform a similar exercise, but separate blocks by the continent that they belong to (Africa, Asia, Europe, North America, Oceania, South America), rather that their income group. Finally, I take this analysis a step further, by dividing the globe into twenty-one subregions around the world which divides the continents into even finer groups. For example Asia is comprised of Central Asia, Eastern Asia, Melanesia, South-Eastern Asia, Southern Asia and Western Asia. I then include about 861 indicator variables representing separate fixed effects for each of the 21 subregions over the period between 1950 and 1990. While the income-group specific time-trends do not affect the baseline estimates by much, the specification using the year-specific time trends for subregions does reduce the size of the coefficient by about half. However, the estimates still remain positive, statistically significant and economically meaningful. This exercise validates the empirical strategy and the baseline results, although it does suggest that a more conservative impact of the Landsat program taking into account region-specific trends.

F. Other Specification Checks:

Finally, in addition to the battery of tests presented above, I also present a number of other robustness checks in Appendix C. Table C.2 presents estimates for different definitions of "low-cloud" in terms of cloud-cover percentage, Table C.3 presents estimates with different start and end years for the panel and Table C.5 presents estimates with standard errors clustered at different spatial groups rather than the Landsat block. Together these robustness checks help to further bolster the validity of the baseline estimates.

4.3 Differential Impact of Landsat Across Firms and Geographies

Having estimated a large and positive impact of Landsat on the discovery of new gold deposits, and having established the robustness of this result to a number of different specifications and tests, I turn to analyzing the differential impact of the Landsat program in order to understand the channels through which publicly-provided information might affect industry.

A. Juniors vs. Seniors: First, I estimate whether Landsat helped both juniors and seniors similarly, or whether it served to narrow or widen the performance differences between these two categories of firms. One mechanism though which publicly-provided maps could boost discovery is by enabling the productivity of smaller and new entrants in the market who might be too capital-constrained to engage in risky, early-stage exploration. This hypothesis predicts that Landsat should be particularly valuable for supporting discoveries from junior exploration firms.

Accordingly, I estimate regressions similar to the baseline specification presented before. However, the dependent variable in these specifications is an indicator variable that is set to one if the discovery is made by either a junior or a senior firm. The estimates of β_1 from such a regression would provide an estimate of the boost to discovery provided to juniors and seniors by the Landsat program, and would allow for a comparison of whether Landsat disproportionately helped one group versus the other.

The estimates from these regressions are presented in Table 5. The estimates suggest that the impact of the Landsat program on juniors is about 0.047, while the impact for seniors is about 0.12. In other words, the total gain from the Landsat program (about 0.16% more discoveries) are split such that smaller firms make 0.04% more discoveries at the block-year level, while seniors capture the remaining 0.12%. Therefore in terms of percentage points it seems like seniors benefit more from the Landsat mapping effort. However, when the previous market-shares of juniors in terms of new discoveries is taken into consideration, this interpretation changes considerably. Specifically, before the Landsat program was launched, juniors made only about 0.008% discoveries in a given block-year on average, while seniors made 0.0694%. This suggests that even though seniors were almost entirely responsible for the new discoveries made in this industry prior to the Landsat program, in mapped regions, juniors made one out of every four discoveries that was reported after Landsat was launched. Given these percent improvements in the likelihood of new discoveries, it seems like the Landsat program helped improve the performance of smaller firms in this industry (juniors) in terms of making new discoveries. In other words, juniors were 5.8x more likely to report a discovery in mapped regions as compared to unmapped regions, while incumbents only benefited by a factor of 1.7x. Therefore, the estimates suggest that even though seniors made a significant portion of new discoveries in mapped regions, their market position eroded considerably, and juniors were able to make considerable gains in performance.[17]

B. The Role of Local Institutions:

Consistent with the notion that Landsat might help capital-constrained juniors to perform exploration activities, we should expect that the impact of the program is the highest where institutional conditions support new entrants. For example, regions with with strong property rights (a particularly important institution in the mining industry) and low levels of corruption are particularly suitable for discoveries for entrants, while for large, industry incumbents, they are less important. Therefore, if allowing capital-constrained entrants to explore is an important channel through which Landsat boosts discoveries, we should expect the baseline effect to be magnified in places with stronger institutional conditions.

In order to test this proposition, I employ three different indicators. First, I rely on the World Bank Classification of nations in five different income categories: High income (OECD), High income (non-OECD), Upper middle income, Lower middle income and Low income. I test for the difference in the main effect of the Landsat program between above-median (High income) and below-median (Upper middle, Lower middle and Low income) countries.

[17] One concern with this interpretation is that seniors might be "outsourcing" their exploration to juniors. In Appendix Table C.10, I use data on joint-ventures to test this idea, and find that while outsourcing could be relevant, a majority of the junior firm discoveries represent discoveries from capital-constrained firms.

Second, I rely on a survey of institutional conditions in the mining industry to obtain a direct measure of local institutional conditions. Specifically, I use data from the Fraser Institute Survey of Mining Companies, that surveys companies about the quality of local institutions that are relevant to mining in different regions around the world. I estimate differential effects for above-median countries (i.e. countries that rank in the top-half of the institutional quality rank distribution) as compared to below-median ones on measures of institutional quality derived from the Fraser institute survey. See section 3.1 D for more details on these data.

Finally, I also employ a historical measure of institutional quality to further triangulate the evidence from the income and survey measures. One limitation with the survey data is that they are measured after the launch of the Landsat program and after new deposits have been discovered, rather than measuring the quality of local institutions when Landsat informationw as first provided. While measures of institutional quality before 1972 are quite difficult to obtain, I rely on the Polity IV dataset described before to classify countries as belonging to democratic versus authoritarian regimes (which the Polity IV dataset scores on a scale of -9 to 9). Importantly, these measures are available for most of the globe before 1972, which allow me to use these historical measures as a complement to income categories and the Fraser Institute survey.

The results from this analysis are presented in Table 6. In each of the three specifications, I present separate estimates for blocks that belong to countries in the top-half of the respective distribution with Column (1) using the data on income categories, Column (2) using data from the Fraser Institute survey and Column (3) using the data from Polity IV. The results indicate quite clearly that the impacts of the program are concentrated in regions in the top-half of the distribution of institutional quality, measured either through income, the Fraser Institute survey or the Polity IV dataset. Specifically, the results indicate that in terms of income categories and the Fraser survey, regions that rank below the median, see no impact of the Landsat information on new discoveries, while when using the pre-1972 Polity IV data, below-median regions do see a positive and statistically significant effect, but the size of the effect is about half as large as experienced by regions above the median.

Overall, the results on the heterogeneous effects of the Landsat program across firm types and regions supports the hypothesis that the Landsat program is more useful for smaller firms and in regions that have supportive local institutions, supporting the interpretation that publicly-provided information eases capital constraints for exploration for new entrants. A striking conclusion from these results is the fact that while Landsat seems to have reduced gaps in performance between larger and smaller firms, it seems to have made high-income regions more productive, exacerbating natural resources inequality between regions.

5 Discussion

This paper estimates the impact of the NASA Landsat satellite mapping program on shaping the discovery of significant new deposits in the gold exploration industry between 1950 and 1990. Using quasi-random gaps in coverage, I find that the availability of mapping information leads to significant new gold discoveries in regions that benefited from the early availability of Landsat information as compared to regions that did not. Quantitatively, the availability of mapping information almost doubles the likelihood of new discoveries in a region after it has been mapped. These effects are magnified for smaller firms and in higher income regions with strong institutions (such as property rights), suggesting that one mechanism through which mapping enables discovery is by supporting the capital-constrained junior sector of the industry. The results speak to the literature on the public-provision of knowledge goods as well as the literature on the role of information goods on regional development.

Back-of-the-envelope calculations suggest that the welfare impact of the Landsat program could be large.[18] For a country the size of the US (with about 3.8 million sq. miles) this translates to additional gold reserves worth about $6.4 billion USD that can be attributed to the information from the Landsat program, the first phase of which cost about $125 million USD. Further, the program also shaped discovery globally, showing how US investments in basic science and technology might have global impacts through the informational channel. The Landsat program in particular, and space science, in general, has routinely been legislated because Congress has found it difficult to

[18]We assume that the marginal impact of additional reserves discovered through the Landsat program on global gold prices was small – perhaps because these discoveries are later exploited and mined at different rates and in response to market conditions, and do not represent an immediate and discrete increase in available gold reserves on the global market.

justify the public expenditure on these efforts (Gabrynowicz, 2005). Even though Landsat data is regularly used in a variety of different sectors (including agriculture, land and water-use planning etc), my estimates suggest that the benefits to the mining sector alone could justify the costs. These calculations could help inform active policy debates about the value of Landsat and similar program.

While the results suggest that public investments in mapping information improve welfare, an important limitation of this study must be acknowledged. Specifically, it is possible that the impact of the mapping information is driven by the substitution of investment in mapped regions as compared to unmapped regions, rather than increases in efficiency on part of firms in the private sector. The results in Table C.1, Panel 4, which estimate limited treatment effects when Landsat maps are uninformative (tree-covered), suggest that purely substitution is unlikely to be at play, however this concern must be acknowledged before the results are taken to policy. Similarly, it is also possible that Landsat merely sped up the rate of discoveries rather than helped make discoveries that would otherwise never have been made. Finally, it is also possible that exploration cost per discovery are significantly higher for smaller firms as compared to larger firms, which would reduce possible welfare benefits of Landsat, which helped shift exploration to this sector of the industry. While substitution across space and time as well as reduced exploration inefficiencies from the junior sector are both reasonable channels that reduce the welfare benefits of Landsat information, more research is needed before concrete welfare statements can be made.

On a broader level, the general finding that public mapping programs, and idiosyncratic yet widespread variations in the quality of mapping information have dramatic impacts on the geography of regional performance seems quite robust and deserving of future research. A related study, Serrato and Wingender (2010) exploits measurement error in Census data, another prominent national mapping effort, and finds that errors in population data have large consequences for regional growth and employment. Further, even in a non-geographic sense, many large-scale publicly funded scientific projects are aimed to providing better scientific maps (such as the Human Genome Map and BRAIN initiative), and anecdotally seem to have shaped and spurred research in their respective fields. This pattern, that publicly-provided mapping information not only describes regions, but also shapes them, seems to have general validity. Therefore opening-up the black-box of mapping as an economic activity, and understanding the antecedents of what features of the landscape get mapped and why, and estimating the consequences of this variation for economic activity seems like a topic ripe for future research.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2000). The colonial origins of comparative development: An empirical investigation. Technical report, National bureau of economic research.

Ashley, R. P., A. F. H. Goetz, L. C. Rowan, and M. J. Abrams (1979). Detection and mapping of hydrothermally altered rocks in the vicinity of the Comstock Lode, Virginia Range, Nevada, using enhanced Landsat images. Technical report, US Geological Survey,.

Azoulay, P., J. S. G. Zivin, D. Li, and B. N. Sampat (2015). Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. Technical report, National Bureau of Economic Research.

Besley, T. J. and R. Burgess (2001). The political economy of government responsiveness: Theory and evidence from India.

Bloom, N., R. Griffith, and J. Van Reenen (2002). Do R&D tax credits work? Evidence from a panel of countries 19791997. *Journal of Public Economics 85*(1), 1–31.

Branch, G. o. C. A. A. a. N. D. C. C. (2009, June). Stages of Mineral Exploration & Development.

Carlson, D. (2014, October). Global Nonferrous Metals Exploration Budget Fell 25% In 2014.

Casaburi, L. and U. Troiano (2016). Ghost-House Busters: The Electoral Response to a Large AntiTax Evasion Program. *The Quarterly Journal of Economics 131*(1), 273–314.

Chevrel, M., M. Courtois, and G. Weill (1981). The SPOT satellite remote sensing mission. *Photogrammetric Engineering and Remote Sensing 47*, 1163–1171.

CHRR and CIESIN - Columbia University (2005). Global Earthquake Hazard Frequency and Distribution.

Cockburn, I. and R. Henderson (1996). Publicprivate interaction in pharmaceutical research. *Proceedings of the National Academy of Sciences 93*(23), 12725–12730.

Czarnitzki, D. and C. Lopes-Bento (2013). Value for money? New microeconometric evidence on public R&D grants in Flanders. *Research Policy 42*(1), 76–89.

David, A. P. and G. Wright (1997). Increasing returns and the genesis of American resource abundance. *Industrial and Corporate Change 6*(2), 203–245.

David, P. A., B. H. Hall, and A. A. Toole (2000). Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy 29*(4), 497–529.

David, P. A., D. Mowery, and W. E. Steinmueller (1992, January). Analysing The Economic Payoffs From Basic Research. *Economics of Innovation and New Technology 2*(1), 73–90.

Davids, C. A. (1986). *Zeewezen en wetenschap: de wetenschap en de ontwikkeling van de navigatietechniek in Nederland tussen 1585 en 1815*. Ph. D. thesis, De Bataafsche Leeuw, Amsterdam.

Dechezleprtre, A., E. Eini, R. Martin, K. T. Nguyen, and J. Van Reenen (2016). Do tax incentives for research increase firm innovation? An RD Design for R&D. Technical report.

Dilley, M., R. Chen, U. Deichmann, A. Lerner-Lam, M. Arnold, J. Agwe, P. Buys, O. Kjekstad, B. Lyon, and G. Yetman (2005). Natural Disaster Hotspots: A Global Risk Analysis. Technical report, The World Bank, Washington, D.C.

Dittmar, J. E. (2011). Information technology and economic change: the impact of the printing press. *The Quarterly Journal of Economics 126*(3), 1133–1172.

Draeger, W. C., T. M. Holm, D. T. Lauer, and R. J. Thompson (1997). The availability of Landsat data: past, present, and future. *Photogrammetric engineering and remote sensing 63*(7), 869–875.

Eisensee, T. and D. Stromberg (2007). News droughts, news floods, and US disaster relief. *The Quarterly Journal of Economics*, 693–728.

Furman, J. L. and S. Stern (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *The American Economic Review*, 1933–1963.

Gabrynowicz, J. I. (2005). Perils of Landsat from Grassroots to Globalization: A Comprehensive Review of US Remote Sensing Law with a Few Thoughts for the Future, The Symposium: Issues in Space Law. *Chicago Journal of International Law 6*, 45–68.

Goldfarb, R. J., T. Baker, B. Dube, D. I. Groves, C. J. Hart, and P. Gosselin (2005). Distribution, character, and genesis of gold deposits in metamorphic terranes. *Economic Geology 100th anniversary volume 40*.

Goolsbee, A. (1998). Does government R&D policy mainly benefit scientists and engineers? Technical report, National bureau of economic research.

Goward, S., T. Arvidson, D. Williams, J. Faundeen, J. Irons, and S. Franks (2006). Historical Record of Landsat Global Coverage. *Photogrammetric Engineering & Remote Sensing 72*(10), 1155–1169.

Hall, B. and J. Van Reenen (2000). How effective are fiscal incentives for R&D? A review of the evidence. *Research Policy 29*(4), 449–469.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *The American Economic Review 102*(2), 994–1028.

Hilson, G. (2002). Harvesting mineral riches: 1000 years of gold mining in Ghana. *Resources Policy 28*(1), 13–26.

Holdings, N. R. (2013). Global 2013 Gold Mine and Deposit Rankings.

Hourian, M. and D. Parkes (2015, February). Federal R&D in the FY 2016 Budget: An Overview.

Howell, S. (2014). Financing constraints as barriers to innovation: Evidence from R&D grants to energy startups. Technical report, Working paper.

Jefferson, R. (2013). Innovation Cartography - 2013 Skoll World Forum on Social Entreprenuership (speech).

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The quarterly journal of economics*, 879–924.

Krohn, M. D., M. J. Abrams, and L. C. Rowan (1978). Discrimination of hydrothermally altered rocks along the Battle Mountain-Eureka, Nevada, mineral belt using Landsat images. Technical report, US Geological Survey,.

Lee, H. A. (1963). *Glacial fans in till from the Kirland Lake fault: a method of gold exploration.* Ottawa, Geological Survey of Canada.

Lerner, J. (2009). *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failedand What to Do About It: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failedand What to Do About It.* Princeton University Press.

Mack, P. E. (1990). *Viewing the Earth: The social construction of the Landsat satellite system.* MIT Press.

Marshall, M. G. and K. Jaggers (2002). Polity IV project: Political regime characteristics and transitions, 1800-2002.

Mazzucato, M. (2015). *The entrepreneurial state: Debunking public vs. private sector myths.* Anthem Press.

McCahon, K. and L. Fredricksen (2014). *Fraser institute annual survey of mining companies.* Fraser Institute.

MODIS Atmosphere Science Team, N. (2005). Cloud Fraction (1 month - Terra/MODIS) | NASA.

Moretti, E., C. Steinwender, and J. Van Reenen (2014). The Intellectual Spoils of War? Defense R&D, Productivity and Spillovers. Technical report, Working Paper, London School of Economics.

Moser, P., A. Voena, and F. Waldinger (2014). German Jewish migrs and US invention. *The American Economic Review 104*(10), 3222–3255.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *The Journal of Political Economy 67*(3), 297–306.

Romer, P. M. (1990, October). Endogenous Technological Change. *Journal of Political Economy 98*(5), S71–S102.

Rowan, L. C. (1975). Application of Satellites to Geologic Exploration: Recent experiments in two spectral regions, the visible and near-infrared and the thermal-infrared, confirm the value of satellite observations for geologic exploration of the earth. *American Scientist*, 393–403.

Rowan, L. C., A. F. Goetz, and R. P. Ashley (1977). Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images. *Geophysics 42*(3), 522–535.

Schodde, R. (2011). Recent trends in gold discovery. In *NewGenGold Conference, Perth*, Volume 2.

Scotchmer, S. (2004). *Innovation and incentives.* MIT press.

Serrato, J. C. S. and P. Wingender (2010). Estimating local fiscal multipliers. *University of California at Berkeley, mimeo*.

Spurr, S. H. (1954). History of forest photogrammetry and aerial mapping. *Photogrammetric Engineering 20*, 551–560.

Squicciarini, M. P. and N. Voigtlander (2015, July). Human Capital and Industrialization: Evidence from the Age of Enlightenment*. *The Quarterly Journal of Economics*, qjv025.

Steinwender, C. (2014). Information Frictions and the Law of One Price:When the States and the Kingdom became United. Technical report, Oesterreichische Nationalbank (Austrian Central Bank).

Stephan, P. E. (2012). *How economics shapes science*, Volume 1. Harvard University Press Cambridge, MA.

Vincent, R. K. (1975). Commercial applications of geological remote sensing. In *Decision and Control including the 14th Symposium on Adaptive Processes, 1975 IEEE Conference on*, pp. 258–263. IEEE.

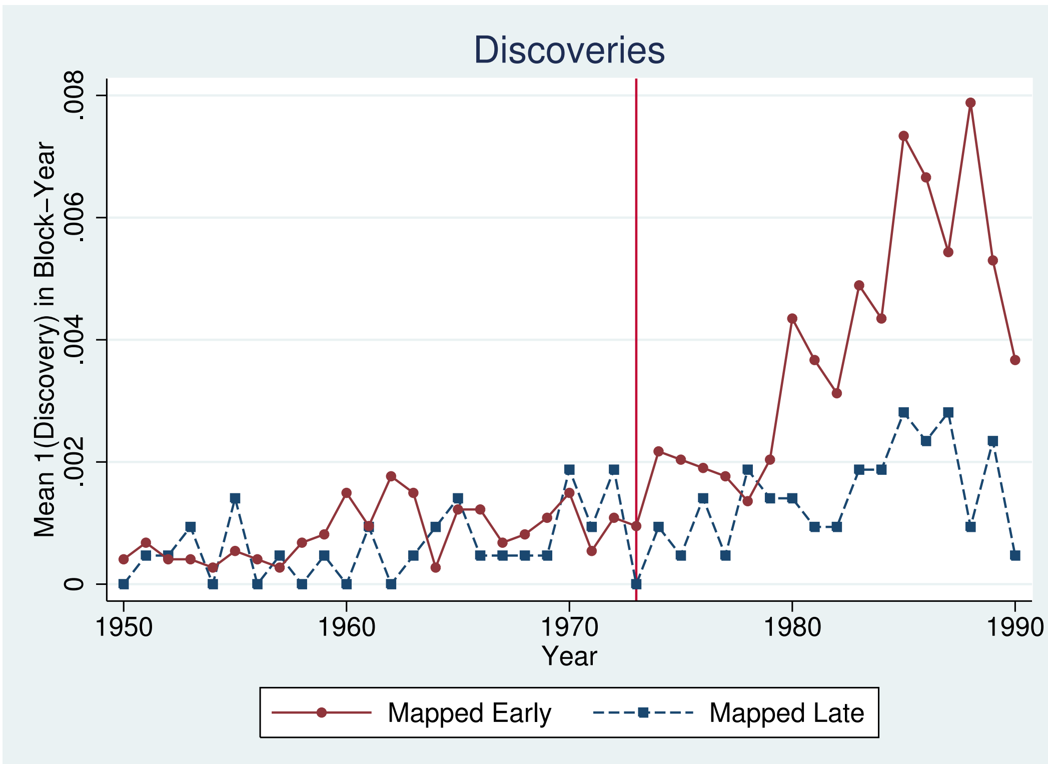
Weatherley, D. K. and R. W. Henley (2013, April). Flash vaporization during earthquakes evidenced by gold deposits. *Nature Geoscience 6*(4), 294–298.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*(1), 1–27.

Wright, B. D. (1983). The economics of invention incentives: Patents, prizes, and research contracts. *The American Economic Review*, 691–707.

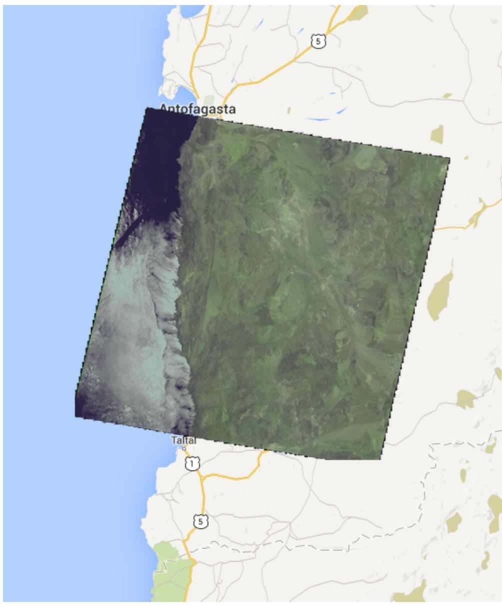
6 Figures and Tables

Figure 1. **Comparing Gold Discovery Rates for Blocks Mapped Early and Late by Landsat**



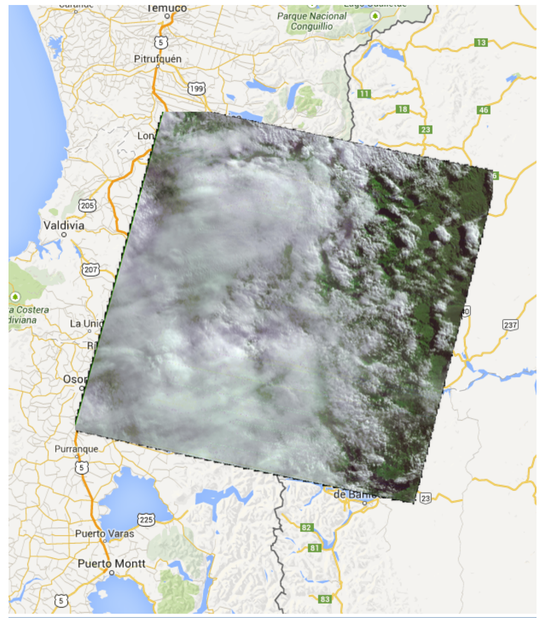
Note: This figure plots the history of global gold discovery over time, using block-year level discovery data. Each block is classified either as a "Mapped Early" coverage block (in red circles), or "Mapped Late" block (in blue squares), depending on whether the year it was first-mapped by the Landsat program was before or after the median first-mapped year of the Landsat program (i.e. before or after 1974). For each block group, average probability of making a discovery is plotted on the y axis and calendar year is on the x axis. The level of observation is block-group by year. For further details on the sample, see the text and data appendix.

Figure 2. **Example of Variation in Landsat Coverage**



(1) Block 25177, Chile
Year mapped (*Post Mapped*) = 1973
Year low-cloud map (*Post Low-Cloud*) = 1973

Outcome: Amax Gold Discovery reported in 1980



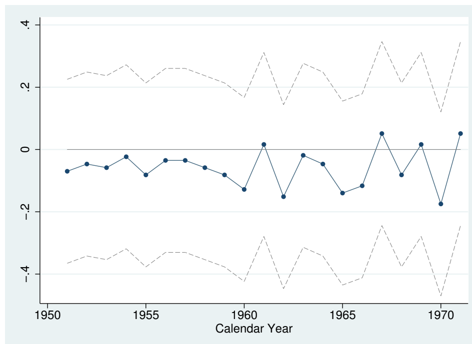
(2) Block 24988, Chile
Year mapped (*Post Mapped*) = 1975
Year low-cloud map (*Post Low-Cloud*) = 1976

Outcome: No discovery reported to date

Note: This figure provides an example that illustrates the research design and data used in the baseline specification using two Landsat blocks in Chile. Figure (1) on the left, shows the best available image for Block 25177 available from the Landsat project by 1983. This map arrived in 1973, relatively early, and happened to be cloud-free. In my data, this block reported a gold discovery by Amax gold in 1980. The best available image for Block 24988 is shown in Figure (2) on the right. This map, the first one for this block, arrived in 1975 (a delay of two years possibly caused due to technical errors in data collection) and happened to be significantly obscured with clouds (upto 39% according to my data) and received a relatively cloud-free map (not depicted above) only in 1976. According to my data, no discovery has been reported in this block by the year 1990.

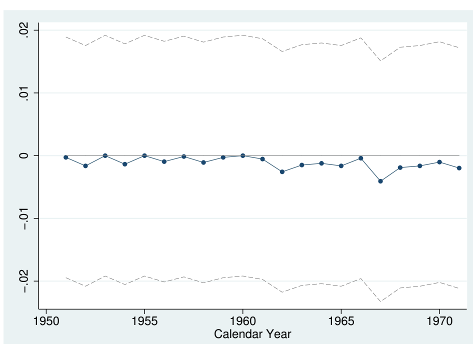
Figure 3. **Comparing Baseline Pre-1972 Characteristics of Blocks Mapped Early and Late by Landsat**

Panel A. Main Outcome: Average Annual 1(Discovery)

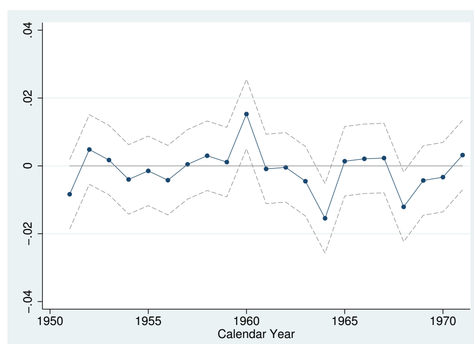


Panel B. Covariates Predicting Prospectivity

Avg. Annual Publications

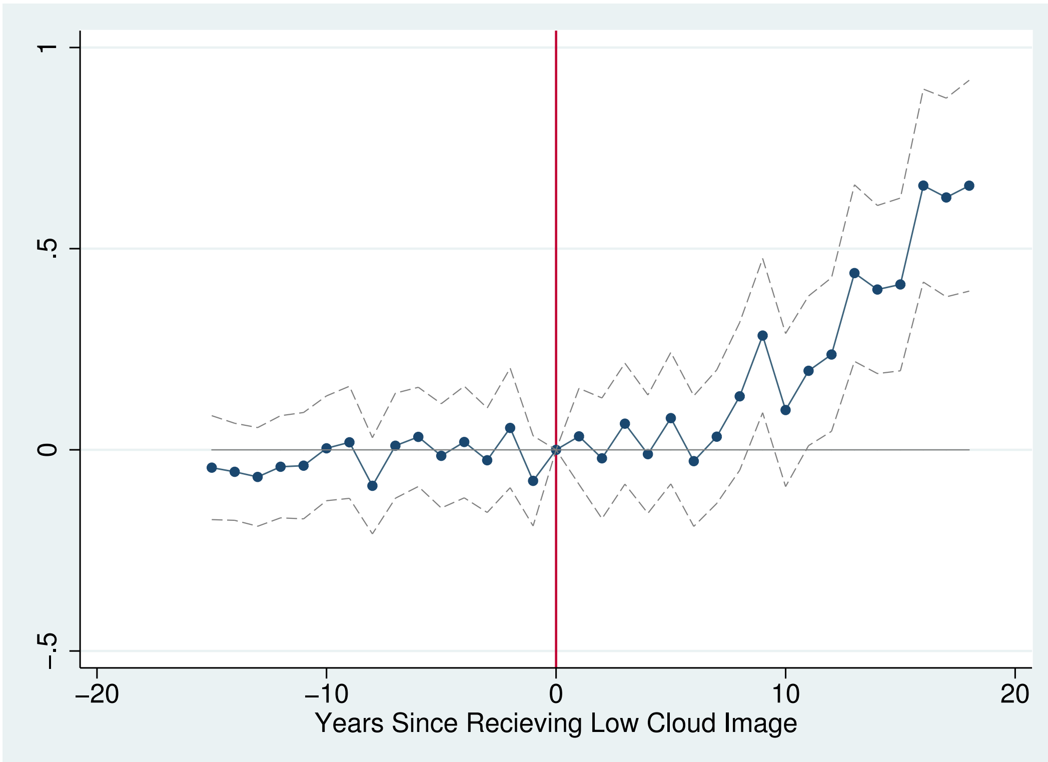


Avg. Annual Earthquakes



Note: This figure explores baseline differences between early and late-mapped blocks. For both panels, difference in means of outcome variable is calculated between blocks mapped early (mapped before 1974) with blocks mapped late (on or after 1974) on a yearly basis to allow a comparison of these variables in levels and trends. The outcome variables are average of indicator variable for discovery in block-year in Panel A, while time-varying outcomes predicting gold prospectivity (average gold-related publications in mining journals and average number of earthquakes) are plotted in Panel B.

Figure 4. **Time Varying Estimates of the Impact of Landsat Intensity on Gold Discovery**



Note: This figure plots estimates (and 95 percent confidence intervals) of β_t from the event study specification specified below. On the x axis is calendar year. This figure is based on block-year observations, the coefficients are estimates from OLS models, the sample includes all block-year discoveries between 1950 and 1990 and the standard errors are robust and clustered at the block level. See the text and data appendix for additional details on variable and data descriptions.

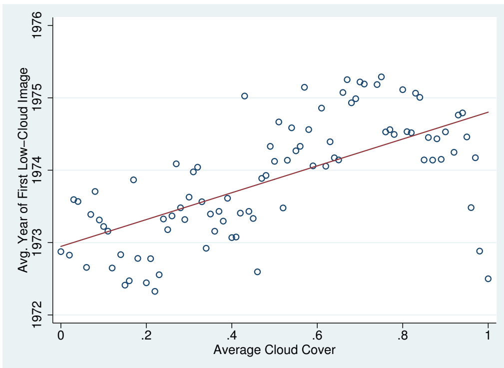
Specification:

$$Y_{it} = \alpha + \Sigma_z \, \beta_t \times 1(z) + \gamma_i + \delta_t + \epsilon_{it}$$

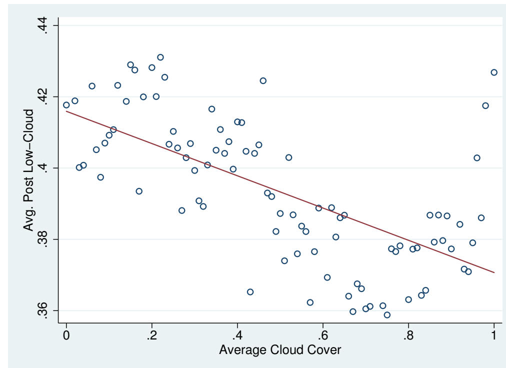
where γ_i and δ_t represent block and time fixed effects respectively for block i and year t. z represents the "lag", or the years relative to a "zero year", which marks the year when a block was first mapped with a low-cloud image (or 1990 if the block was never mapped).

Figure 5. **Binned Scatterplots Testing the Cloud Cover Instrument**

Panel A: First Stage: Cloud Cover and Image Timing

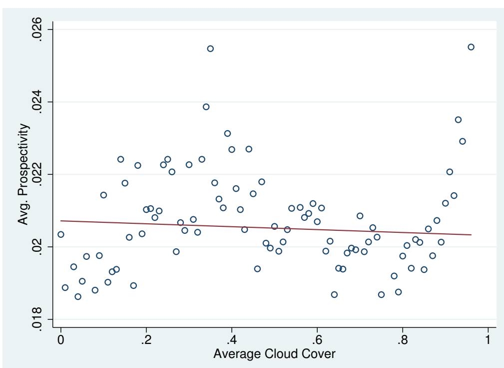


(1) Avg. Year of First Low-Cloud Image

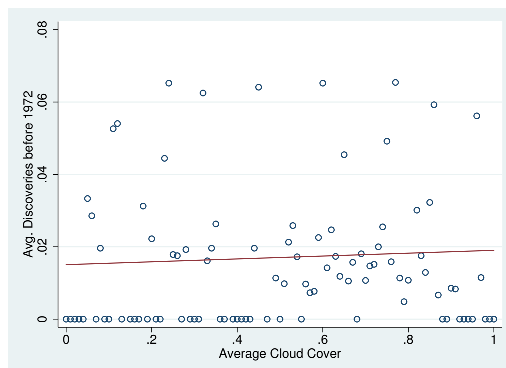


(2) Avg. Post Low-Cloud Indicator

Panel B: Exclusion Restriction: Cloud Cover and Correlates of Gold Discovery



(1) Avg. Predicted Prospectivity



(2) Avg. Discoveries before 1972

Note: This figure plots the relationship between average annual cloud cover and the timing of Landsat images (Panel A) and between the average annual cloud cover and correlates of gold discovery at the block-level (Panel B). For all four charts, blocks are binned by the level of average annual cloud cover rounded to two decimal digits, and mean value of the variable on the y-axis is calculated. Panel A records the first-stage relationship between the cloud cover instrument and the endogenous variable. Outcome variable in Panel A, Figure 1 is the year in which the block recieved a low cloud cover image, while the variable in Panel A, Figure 2 is the average of the indicator variable for whether a low-cloud image is available. Panel B, tests the correlates of cloud cover with other variables that could affect gold discovery. I predict a "prospectivity" score for a given block as a function of gold-mining publications (before 1972) and earthquake-risk index based on the geology of the region. A mean value of this score is graphed on the y-axis. Similarly, Panel B, Figure 2 plots the average number of discoveries before 1972 on the y-axis.

Table 1. **Summary Statistics**

Panel A – Block - Year Level

	Mean	SD	Median	Min	Max
Outcome					
Any Discovery (%)	0.188	4.33	0.000	0	100
Any Junior Disc. (%)	0.038	1.94	0.000	0	100
Landsat Coverage					
Post Mapped	0.409	0.49	0.000	0	1
Post Low-Cloud	0.381	0.49	0.000	0	1
Block-year Covariates					
Publications	0.009	0.40	0.000	0	76
Num. Earthquakes	0.017	0.21	0.000	0	20

Panel B – Block Level

	Mean	SD	Median	Min	Max
Outcome					
Total Discoveries	0.083	0.52	0.000	0	16
Total Junior-led Disc.	0.017	0.18	0.000	0	7
1(Ever Discovered)%	4.846	21.47	0.000	0	100
1(Discovered post-1972)%	3.940	19.45	0.000	0	100
Landsat Coverage					
Year First Mapped	1973.222	3.58	1972.000	1972	1990
Year First Low-Cloud	1974.368	5.19	1972.000	1972	1990
Block Covariates					
Tree Cover(%)	0.208	0.41	0.000	0	1
Avg. Annual Cloud Cover	0.627	0.24	0.676	0	1
Predicted Prospectivity Score	1.896	0.85	1.577	2	15
Publications(pre-72)	0.025	0.66	0.000	0	47
Earthquake Hazard Index	0.628	1.67	0.000	0	10

Note: Observations at the Block – Year level for Panel A and at the Block level for Panel B. A "block" is a Landsat image or scene as defined by the Worldwide Reference System (WRS-1) which divides the planet into blocks of approximately 180km X 180km. I include all blocks that cover the earth's landmass excluding blocks that are comprised purely of water bodies (as well as Antarctica and Greenland), resulting in a total of 9493 blocks in my sample. The period for the analysis is 1950 – 1990. See text for data and variable descriptions.

Table 2. **Cross-sectional Comparison of Blocks by Landsat Coverage**

Panel A – Comparison by Landsat Timing

	(1) mapped-early	(2) mapped-late	(3) diff	(4) p-val
Discoveries (pre-72)	0.020	0.015	0.006	0.22
Prospectivity Score	1.925	1.796	0.129	0.00
Publications (pre-72)	0.032	0.001	0.031	0.05
Earthquake Hazard	0.683	0.440	0.242	0.00

Panel B – Comparison by Avg. Annual Cloud Cover

	(1) Low Cloud	(2) High Cloud	(3) diff	(4) p-val
Discoveries (pre-72)	0.018	0.020	-0.001	0.76
Prospectivity Score	1.893	1.899	-0.007	0.70
Publications (pre-72)	0.014	0.036	-0.022	0.10
Earthquake Hazard	0.628	0.629	-0.001	0.98

Note: This table compares cross-sectional differences between blocks in terms of four covariates for two different subsamples. Panel A compares blocks mapped early (before 1974) and late (in or after 1974) with low-cloud images by the Landsat program. Panel B compares blocks with low amount of cloudiness (below the median value of 67%) with blocks with high amount of cloudiness. Column (3) is the estimate for the difference in means, and column (4) is the p-value for the t-test that the difference in means is significantly different than zero. Discoveries(pre-72) is the total number of discoveries made in a block before 1972. Prospectivity Score is a score for the predicited prospectivity of a block based on a regression on pre-1972 discoveries on block-level covariates that predict prospectivity. Publications (pre-72) denotes the total number of gold-mining related publications about a certain block published before 1972, while Earthquake Hazard is an estimate of how prone a certain block is to earthquakes.

Table 3. **Baseline Estimates for the Impact of Landsat on Gold Discovery**

	Any Discovery	Any Discovery	Any Discovery	Any Discovery
Post Mapped	0.251***		0.152***	
	(0.0265)		(0.0294)	
Post Low-Cloud		0.267***		0.164***
		(0.0276)		(0.0274)
Block FE	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	389213	389213	389213	389213

*+:p<0.15; *:p<0.10; **:p<0.05; ***:p<0.01*
Standard errors clustered at block-level shown in parentheses.

Specification: $Y_{it} = \alpha + \beta_1 \times Post_{it} + \gamma_i + \delta_t + \epsilon_{it}$ where γ_i and δ_t represent block and time fixed effects respectively for block i and year t.

Note: Block-year level observations. All estimates are from OLS models. The sample includes all block-years from 1950 to 1990 (9493 blocks for 41 years implies 389,213 block-year observations). Post Mapped: 0/1 =1 for a block-year after the first image has been received and Post Low-Cloud: 0/1 =1 for block-year after the first low-cloud image (lower than 30% cloud cover) has been received. 1(Discovery): 0/1 =1 if a discovery is reported in a block-year. See text and appendix for data and variable descriptions.

Table 4. **Instrumental-variables Estimates for the Impact of Landsat on Gold Discovery**

	Post Low-Cloud	Any Discovery
Avg. Cloud Cover X 1(IsOperational)	0.104***	
	(0.00525)	
Post Low-Cloud		1.126**
		(0.484)
Block FE	Yes	Yes
Year FE	Yes	Yes
N	389213	389213
F-Stat	394.03	

*+:p<0.15; *:p<0.10; **:p<0.05; ***:p<0.01*
Standard errors clustered at block-level shown in parentheses.

Note: This table presents instrumental variable estimates relating discovery and discovery-value to the indicator variable for whether a low-cloud image was obtained at the block-year level (Post Low-Cloud), instrumented by a measure of avg. annual cloud cover at the block level (Avg. Cloud Cover) interacted with a dummy variable for whether the program is operational in the block's region (1(IsOperational)). Block-year level observations. All estimates are from OLS models and include block and year fixed effects. The sample includes all block-years from 1950 to 1990 (9493 blocks for 41 years implies 389,213 block-year observations). See text and appendix for data and variable descriptions.

Table 5. **Impact of Landsat on Gold Discovery for Different Types of Firms**

	1(Junior)	1(Junior)	1(Senior)	1(Senior)
Post Mapped	0.0288***		0.127***	
	(0.00563)		(0.0285)	
Post Low-Cloud		0.0472***		0.121***
		(0.00651)		(0.0260)
Percent Gain	**355.68%**	**583%**	**182.39%**	**174.95%**
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	389213	389213	389213	389213

*+:p<0.15; *:p<0.10; **:p<0.05; ***:p<0.01*
Standard errors clustered at block-level shown in parentheses.

Specification: $Y_{it} = \alpha + \beta_1 \times Post_{it} + \gamma_i + \delta_t + \epsilon_{it}$ where γ_i and δ_t represent block and time fixed effects respectively for block i and year t.

Note: Block-year level observations. All estimates are from OLS models. The sample includes all block-years from 1950 to 1990 (9493 blocks for 41 years implies 389,213 block-year observations). Post Mapped: 0/1 =1 for a block-year after the first image has been received and Post Low-Cloud: 0/1 =1 for block-year after the first low-cloud image (lower than 30% cloud cover) has been received. 1(Junior): 0/1 =1 if a discovery is reported in a block-year by a junior mining firm and 1(Senior): 0/1=1 if a discovery is reported in a block-year by an non-junior entity. See text and appendix for data and variable descriptions.

Table 6. **Impact of Landsat on Gold Discovery by Quality of Local Institutions**

	Income	Inst. (Survey)	Inst. (Polity IV)
Post Low-Cloud	0.0398	0.0116	0.0943***
	(0.0288)	(0.0277)	(0.0329)
Post Low-Cloud X Above-Median	0.352***	0.437***	0.117***
	(0.0540)	(0.0563)	(0.0420)
Block FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	389213	389213	389213

*+:p<0.15; *:p<0.10; **:p<0.05; ***:p<0.01*
Standard errors clustered at block-level shown in parentheses.

Specification:

$$Y_{it} = \alpha + \beta_1 \times Post_{it} + \beta_2 \times Post_{it} \times 1(High_i) + \gamma_i + \delta_t + \epsilon_{it}$$

Note: Block-year level observations. All estimates are from OLS models. The sample includes all block-years from 1950 to 1990 (9493 blocks for 41 years implies 389,213 block-year observations). Post Low-Cloud: 0/1 =1 for a block-year after the first low-cloud image (lower than 30% cloud cover) has been received. 1(High): 0/1=1 for blocks that are above the median in terms of rank on the dimension under study. To qualify as "Above-Median", In column 1, blocks must belong to high-income countries (OECD as well as non-OECD) according to the World Bank classification of nations, in column 2, blocks must belong to countries in the top-half the "policy rank" distribution according to the Fraser Institute Mining Survey and in column 3, blocks must belong to countries with a positive Polity score before 1972 according to the Polity IV dataset (indicating a democratic rather than authoritative orientation). See text and appendix for data and variable descriptions.

7 Appendices (For Online Publication Only)

Appendix A: Data description

This appendix describes in additional detail the data sets used in the analysis.

A1. Landsat Coverage Data

It is useful to review some technical details of the Landsat satellite program before understanding how the data on Landsat coverage are generated.

Landsat Program: The Landsat program is a forty year-old program to collect imagery of the earth's surface. There have in total been seven successful Landsat satellite launches including Landsat 1, 2 and 3 which form the first generation of satellites launched in the 1972, 1975 and 1978 respectively. These first generation satellites had a similar technical design and are the focus of this paper. Each operated at an orbit of about 900 km above the earth's surface, took images of "moderate resolution" covering an area of approximately 185km X 185km in each image. Each satellite orbited the earth once every 18 days, and consequently was designed to collect repeat images of the earth's surface over this interval. Each satellite carried the "Multispectral Scanner System" (MSS sensor) that captured information in the spectral resolution of $0.5 - 1.1$ μm (Landsat Data Users Handbook). Because images were taken using the MSS rather than a standard optical camera, different bands of information were captured including the visible IR and reflected near-IR portions of the spectrum, and all of this data was available for analysis. Note that Landsat-1 also contained another sensor, the "Return Beam Vidicon" (RBV) sensor, which proved to be a subsidiary sensor and provided very little data.[19] I exclude the RBV sensor from the analysis and focus on the MSS only. The Landsat system operated under the "Worldwide Reference System" (WRS) that is a referencing system to identify different locations around the earth and their corresponding image in the Landsat system. By my calculations, about 15,000 of these Landsat image locations intersect land features on the earth, and these 15,000 185kmx185km "blocks" form the sample for my analysis.

Calculating Landsat Intensity by block: In 1973, The US Geological Survey (USGS) constructed a facility near Sioux Falls, South Dakota known as the Earth Resources Observation and Science (EROS) data center to archive and distribute Landsat imagery. This data center is the main repository of Landsat information for follow-on use. In order to quantify variation in the availability of imagery it is necessary to study these archives and arrive at estimates of Landsat holdings at a given point in time for a given block, a non-trivial exercise, given both the size of the data holdings and the difficulties in accessing the data. I rely on Goward et al. (2006), a study conducted at the behest of the Advisory Committee to the USGS National Satelite Land Remote

[19] see https://lta.cr.usgs.gov/rbv.html.

Sensing Data Archive to produce estimates of Landsat historical holdings at different points in time. This study analyzed all available data at the EROS center and produced maps that visualized geographical variation in Landsat holdings on a yearly basis and also provided estimates of cloud cover for each block-year. These data are publicly available and I downloaded them from http://edcftp.cr.usgs.gov/pub//data/richness/ on Nov 13, 2014. Specifically I focus on the `mss1` files available (from `mss1972.tar.gz` upto `mss1983.tar.gz`) as the raw maps used to calculate my Landsat intensity scores. Each of these files provides a shape file for the corresponding year, that show the number of images collected for each block as well as measures of cloud cover in 10% increments. Using these block-year level observations of intensity and cloud cover, I calculate overall measures of intensity and cloud cover.

Note that while the EROS center was the major repository of Landsat imagery charged with providing data without discrimination, globally and at a reasonable cost – it was not the only source of Landsat data. There were a few countries that collected local Landsat data through "International Cooperator" (IC) stations, and some US departments maintained their personal repository of Landsat data. For the purposes of this paper, I am unable to survey these data and rely on the estimates from Goward et al. (2006) for the analysis in this paper.

A2. MinEx Consulting Discovery Database

There exists no canonical database that tracks global mineral discoveries. In this paper I use a proprietary database developed by MinEx Consulting, Australia to track gold discoveries. These data have been compiled manually over many years by Mr. Richard Schodde of MinEx consulting. These data are based on information sourced from company annual reports, press releases, NR 43-101 disclosure documents under Canadian law, technical and trade journals (like Economic Geology, Northern Miner and Mining Journal), government files from various national geological surveys and personal communications with key people in the industry. These data were made available for the current research project under a non-disclosure agreement with MinEx consulting and are not available for redistribution given their commercial value.

Coverage: The data was compiled from MinEx's master database which contains information on over 55,000 mineral deposits across a wide range of metals. A large number of these deposits are smaller than "Minor" – and as such are of limited commercial interest. It is unlikely that the database has 100% coverage, because a large number of deposits (especially of smaller sizes) are not reported systematically and many companies and countries do not provide full breakdowns of their current inventory. Notwithstanding these issues, according to MinEx estimates that "its database (including information on discovery date) for Gold and Base Metals captures at least 99% of all giant-sized deposits and 93% of the major deposits and 65% of the moderate deposits" across all minerals. Coverage for larger deposits for gold is estimated to be significantly better than this baseline estimate.

Additional Notes: It is important to note a few important aspects of these data. First, while I analyze "gold" deposits, deposits are often made up of more than one mineral. Gold for example, is often found along with Copper. I only include deposits where the "primary metal" has been identified as Gold according to the MinEx data. MinEx makes this evaluation based on the economic value of the different mineral deposits reported at a given location.

Second, the discovery year refers to when the deposit was recognized as having significant value. This is usually set as the date of the first economic drill intersection. It should be noted that a review of the discovery history of the deposit may show that there were small-scale workings on the site. For purposes of these data, if there is a order-of-magnitude step change in the known endowment of the deposit (i.e. from 100 koz to >1 Moz of Gold, the date of the upgrade is viewed as the discovery date for the main deposit.

Appendix B: Back-of-the-Envelope Welfare Calculation

It is quite difficult to calculate the exact contribution of the Landsat program to welfare. Such a calculation would involve the total general equilibrium impact of the program on a number of different margins including (a) additional private sector surplus, (b) additional consumer surplus for end users and (c) costs of the program. The qualitative literature on Landsat (Mack, 1990) has documented the large number of applications of Landsat information in a wide variety of diffent sectors including agriculture, land-use and urban planning, environmental and geological research, forestry, hydrology, transporation etc. Evaluating the general welfare contributions of the program to all of these different sectors though desirable, is beyond the scope of this calculation.

Instead, I will make a number of (perhaps restrictive) assumptions to arrive at one reasonable estimate for the impact of the Landsat program. Specifically, I will focus on the value of the Landsat program to the gold exploration industry between 1972 and 1990. This will help calculate one lower bound of the value of the program. Further, I will assume that the discovery of the program has few general equilibrium impacts, especially on gold price, which will be an important determinant of the overall profits to the industry. Further, the increased availability of gold as a natural resource due to Landsat could have had implications for consumers, especially because gold is a common material used in technological applications like electronics and computer chips—I will ignore these consumer effects and will focus instead on the value for the firms. Finally, there are significant costs involved in gold discovery and exploration. For the purposes of this analysis, I will focus on additional revenues from new gold discoevries and ignore additional costs, mostly because of a lack of data on search costs in my setting.

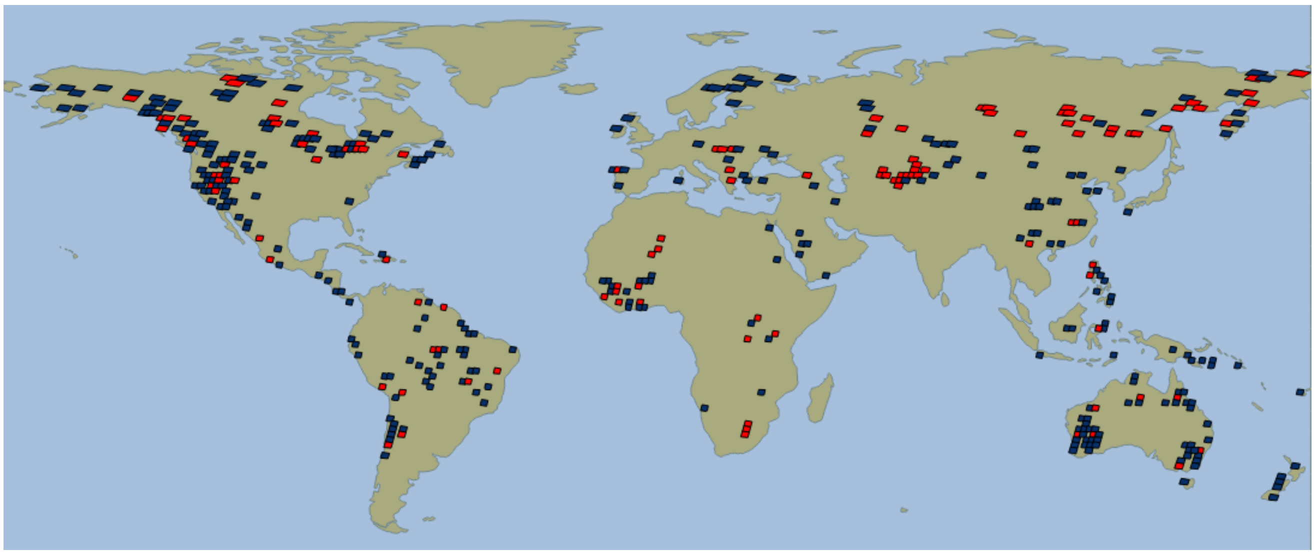
With these caveats, a back-of-the-envelope calculation of the value of the Landsat program would proceed as follows. My estimates from 3 suggest that in blocks that benefited from the availability of Landsat mapping information, the probability of new discoveries rose by about 0.164 (column 4), or about 0.00164 additional discoveries per block-year. This translates to about 0.0246 additional discoveries over a fifteen year period. In my data, of all discoveries made between 1950 and 1990, the average discovery size is about 1.8 Moz, and therefore each mapped is likely to discover about

0.04428 Moz of additional gold reserves due to the Landsat program. From a block's point of view, this translates to about \$17.7 million USD assuming a gold price of \$400 per ounce (which was the price of gold in the 1980s). At the present rate of about \$1300 per ounce, this translates to additional deposits worth about \$57.5 million USD. For a country the size of the US (with about 3.8 million sq. miles) this translates to additional gold reserves worth about \$6.4 billion USD that can be attributed to the information from the Landsat program (at a gold price of \$400 per ounce).

The cost of the Landsat program were relatively modest in comparison. The first generation of the Landsat program was estimated to have cost about \$125 million USD (Mack, 1990). Therefore, even discounting the contributions of the Landsat program to numerous other sectors, the value of the Landsat program to the gold exploration industry alone seems to have justified its costs. Even if exploration costs were estimated to be about 50% of total value of reserves, this conclusion would not change significantly. It is quite clear from such a back-of-the-envelope calculation that the Landsat program created enormous value for the gold exploration industry.

Appendix C

Figure C.1. **Blocks with Gold Discoveries**

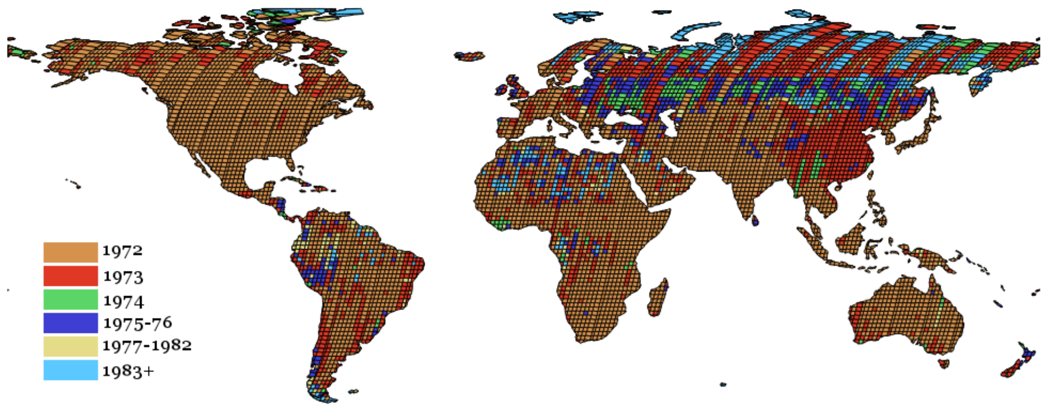


Key– 1958-1973: Red and 1973-1988: Blue

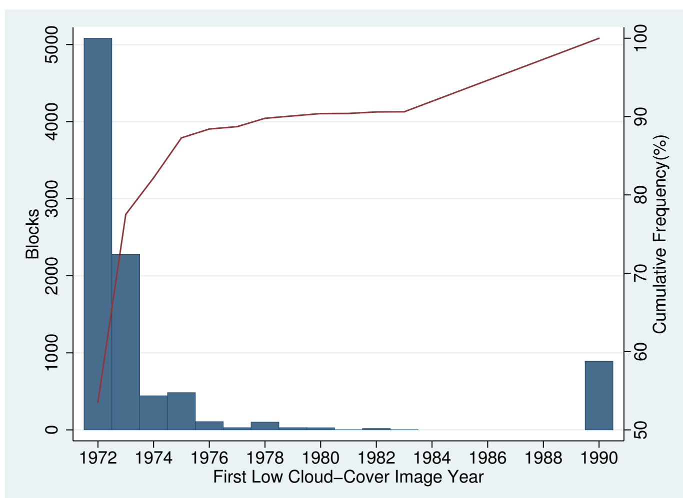
Note: This map plots blocks that reported gold discoveries of significant size as reported by the MinEx data. The blocks are color coded by the first year that discovery was reported since 1958; red if this year was before 1973 and blue if it was after. Note that blocks can (and sometimes do) report discoveries in multiple years, in which case, they are color coded by the first year in which the discovery was reported.

Figure C.2. **Variation in Mapping Coverage**

Panel A: Landsat "Blocks" and Years First Mapped by Landsat



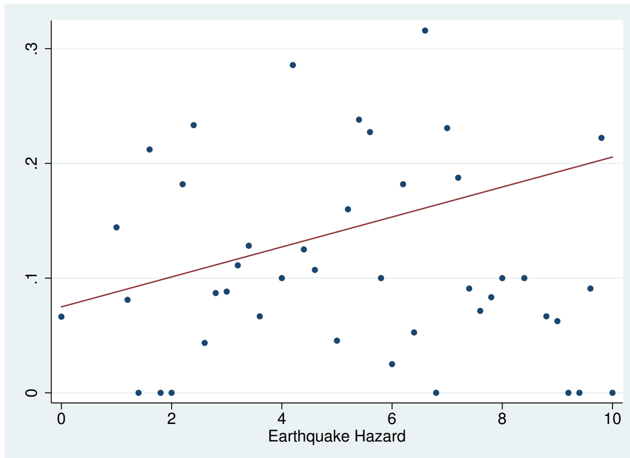
Panel B: Time-series Variation in Landsat Coverage (First Low-Cloud Image Year)



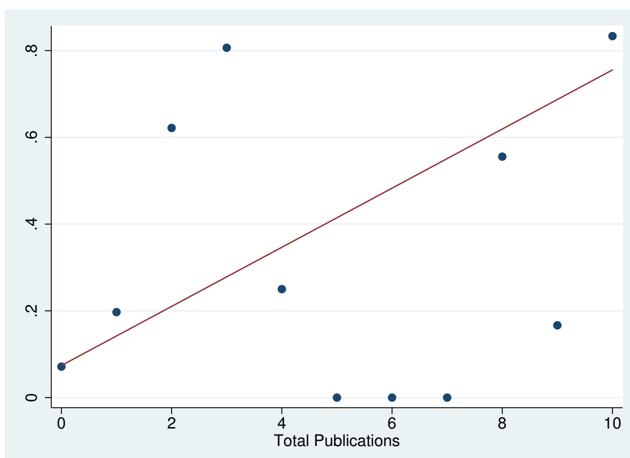
Note: This figure illustrates Landsat blocks and the variation in their mapping over time exploited in this paper. Panel A shows the location of each Landsat block, and the color represents the year in which these blocks were first mapped by the Landsat program. For blocks that were not mapped by the first phase of the Landsat program are represented as "1983+" in light blue. Panel B plots a histogram for the year in which blocks were first mapped with a low cloud cover images. The frequency counts of the blocks are on the left y-axis and cumulative frequency in percent is represented on the right y-axis. Blocks shown to have been mapped in 1990, were blocks that were not mapped by the first phase of the Landsat program, but were mapped later by following generations of satellites.

Figure C.3. **Testing the Validity of the Prospectivity Score Measure**

Panel A: Earthquake Propensity and Gold Discovery



Panel B: Geological Research and Gold Discovery



Note: This figure outlines binned scatterplots outlining the relationship between gold discovery and predictors of gold discovery used in the gold propensity score index. In Panel A, the variable on the x-axis is earthquake hazard index at the block level, while in the Panel B, the variable on the x-axis is the total number of publications for a given block. For both panels, the y-axis is the total number of discoveries between 1950 and 1990. For Panel A, blocks are grouped at values at equal intervals (rounded to the closest 0.2) and block-groups with over 0.4 average findings are omitted, while for Panel B, all blocks with more than 10 publications are normalized to have 10 publications. The fitted lines are weighted by bin size for both panels.

Table C.1. **Robustness Checks – Subsample analysis**

	1(Discovery)	1(Discovery)
Panel A : Excluding USA–		
N=389336		
Post Mapped	0.136***	
	(0.0292)	
Post Low-Cloud		0.141***
		(0.0270)
Panel B: No USA, Can, Aus		
N=291510		
Post Mapped	0.106***	
	(0.0305)	
Post Low-Cloud		0.0929***
		(0.0277)
Panel C: No USSR		
N=383719		
Post Mapped	0.0862*	
	(0.0508)	
Post Low-Cloud		0.101**
		(0.0421)
Panel D: No pre-72 disc.		
N=383719		
Post Mapped	0.130***	
	(0.0251)	
Post Low-Cloud		0.157***
		(0.0243)
Panel E : Tree-Cover Only		
N=80975		
Post Mapped	-0.0302	
	(0.137)	
Post Low-Cloud		0.0492
		(0.103)
Block FE	Yes	Yes
Year FE	Yes	Yes

Note: This table presents estimates from baseline DD specification for different subsamples of the data. Panel A excludes USA blocks, Panel B excludes blocks from USA, Canada and Australia, Panel C excludes any blocks that reported at least one discovery in the period 1950-1972 and Panel D only includes blocks that have substantial tree cover. Tree cover is coded using codes 1-4 from the GLC2000 dataset indicating all

Table C.2. **Robustness Checks – Different Cutoffs for Low Cloud Cover**

	10 pct.	20 pct.	40 pct.	50 pct.
Post Low-Cloud	0.156***	0.148***	0.180***	0.165***
	(0.0287)	(0.0282)	(0.0262)	(0.0277)
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	389213	389213	389213	389213

Note: This table presents estimates that use different cutoff points to calculate an image with low cloud cover. In the main specification, any image below 30 percent cloud cover is considered a low cloud image. This table evaluates the baseline regression with a block considered to be mapped with a low cloud image if an image was obtained with cloud cover values below 10%, 20%, 40% or 50% in Columns 1-4 above.

Table C.3. **Robustness Checks – Different Panel Lengths**

	1950-1990	1951-1989	1952-1988	1953-1987	1954-1986	1955-1985
Post Low-Cloud	0.163***	0.160***	0.161***	0.145***	0.131***	0.112***
	(0.0274)	(0.0281)	(0.0278)	(0.0284)	(0.0288)	(0.0293)
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	389213	370227	351241	332255	313269	294283

Note: This table presents estimates for the Landsat program by differing length of the panel. In each of the columns data is only included for years 1950-90, 1951-89, 1952-88, 1953-87, 1954-86 and 1955-90 respectively.

Table C.4. **Robustness Checks – Excluding Blocks Treated Early**

	Excluding 1972	Excluding 1972	Excluding 1972-1973	Excluding 1972-1973
Post Mapped	0.133***		0.0825+	
	(0.0398)		(0.0507)	
Post Low-Cloud		0.115***		0.0902*
		(0.0339)		(0.0473)
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	137268	180851	50758	87494

Note: This table presents estimates for the Landsat program by excluding blocks that were treated in the first year, or the first two years of the program. A majority of the blocks were treated in this period, so this analysis restricts attention only within blocks that experienced a delay of at least two years or more. Columns (1) and (3) exclude 64.7% (6145) and 86.7% (8255) blocks that were mapped within the first or first two years of the program, respectively while columns (2) and (4) exclude 53.5% (5082) and 77.5% (7359) blocks which were mapped with low-cloud maps within the first, or first two years of the progam respectively.

Table C.5. **Robustness Checks – Alternate Spatial Clustering**

	Any Discovery	Any Discovery	Any Discovery	Any Discovery
Post Mapped	0.152***		0.152***	
	(0.0315)		(0.0349)	
Post Low-Cloud		0.164***		0.164***
		(0.0306)		(0.0350)
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Clustering Block Size	2x3	2x3	5x6	5x6
N	389213	389213	389213	389213

Note: This table presents estimates for the Landsat program similar to the specification in the baseline specification except with different assumptions for the clustering of standard errors. In all specifications, standard errors are clustered at "block groups", which are sets of blocks larger than any given block. In columns (1) and (2), the size of a block group is 2x3 blocks (i.e. 6 blocks in any given block group), while in columns (3) and (4), the size of a block group is 5x6 (i.e. 30 blocks in any given block group). Standard errors are clustered at this large block-group level rather than at the block level.

Table C.6. **Robustness Checks – Placebo Treatment Year**

	Any Discovery	Any Discovery
Post Low-Cloud	-0.0521	-0.0362
	(0.0424)	(0.0395)
Block FE	No	Yes
Year FE	Yes	Yes
N	389213	389213

Note: This table presents estimates for the Landsat program similar to the specification in the baseline specification except the "Year Low-Cloud" is replaced by a random integer between 1972 and 1988 for each block, effectively randomizing the treatment year at the block-level. The other variables are calculated as usual.

Table C.7. **Robustness Checks – Negative Binomial Specification**

	Any Discovery	Any Discovery	Any Discovery	Any Discovery
isfind				
Post Mapped	2.640***		1.626**	
	(0.581)		(0.661)	
Post Low-Cloud		1.526***		0.609*
		(0.329)		(0.358)
Block FE	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	389213	389213	18860	18860

Note: This table presents estimates for the Landsat program similar to the specification in the baseline specification, except the estimates are produced from negative-binomial regressions, rather than OLS models. Exponentiated versions of the above coefficients provide estimates of the treament elasticities.

Table C.8. **Robustness Checks – Differential Time Trends by Region Types**

	Any Discovery	Any Discovery	Any Discovery
Post Low-Cloud	0.171***	0.0713**	0.0729**
	(0.0260)	(0.0292)	(0.0307)
Block FE	No	Yes	Yes
Time FE	Income Grp. X Year	Continent X Year	Subregion X Year
N	389213	389213	389213

Note: This table presents estimates for the Landsat program similar to the specification in the baseline specification, except instead of a common year fixed effect across regions, time fixed effects are estimated using region-specific time trends. Column (1) includes separate time trends for five different income groups interacted with year dummies, Column (2) includes time trends for six continents (Africa, Asia, Europe, North America, South America, Oceania) interacted with year dummies while Column (3) includes separate time trends for 21 separate subregions (for example, "south-eastern asia" or "western europe") interacted with year dummies.

Table C.9. **Cross-Sectional Specification: Are Delays Associated with Lower Rate of Discoveries?**

	Any Disc. (72-90)	Any Disc. (72-90)	Any Disc. (72-90)	Any Disc. (72-90)
Delay (Years)	-0.231***	-0.144**	-0.148***	-0.105***
	(0.0206)	(0.0423)	(0.0479)	(0.0330)
Fixed Effects	None	Continent FE	Subregion FE	Block-Group FE
adj. R^2	0.00370	0.0143	0.0168	0.0208
N	9493	9493	9493	9493

Note: This table presents an alternate cross-sectional specification to understand the impact between Landsat mapping delays and gold discovery. Specifically, I estimate an equation of the form $Y_i = \alpha + \beta_1 \times Delay_i + \gamma_i + \epsilon_i$, where Y_i the main outcome variable is an indicator for whether any discovery was made in a given block i after the launch of the Landsat program in 1972 till 1990, $Delay_i$ is the total difference between the year in which a block was mapped with a low cloud image and 1972, γ_i represents spatial fixed effects. In Column (2), I include six separate continent fixed effects (North America, South America, Asia, Africa, Europe and Oceania). In Column (3), I include 21 separate subregion fixed effects, where subregions include Central Asia, Northen Africa, Western Europe, Carribean etc. In Column (4), I include separate "block-group" fixed effects where I divide all possible blocks into thirty-four large block-groups. Standard errors are clustered at the block level in all specifications. Standard errors are clustered at the same level as the fixed effects, except in Column (1) where robust standard errors are reported.

Table C.10. **Robustness Check – Junior and Senior Discovery Accounting for Joint Ventures**

	1(Junior)	1(Junior)	1(Senior)	1(Senior)
Post Mapped	0.0227***		0.125***	
	(0.00523)		(0.0283)	
Post Low-Cloud		0.0351***		0.119***
		(0.00566)		(0.0259)
Percent Gain	**294.64%**	**456.15%**	**182%**	**173.32%**
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	389213	389213	389213	389213

Note: This table presents estimates for the Landsat program accounting for joint discoveries. In some cases, new discoveries are announced by more than one firm, often a junior and a senior. While the main specification codes firm type of the majority stakeholder in the project (either a junior or a senior), for this regression, all joint-ventures are assumed to be senior-led discoveries, while junior-led discoveries include only those projects where only one junior firm was involved in the discovery.