THE PRIVATE IMPACT OF PUBLIC MAPS— LANDSAT SATELLITE IMAGERY AND GOLD EXPLORATION

Abhishek Nagaraj[*]
nagaraj@mit.edu

JOB MARKET PAPER

November 6, 2015

Abstract

For centuries, the availability of maps of under-explored geographies has provided new opportunities for innovators, and yet mapping as a channel to enable discovery has been rarely examined. To shed light on this topic, I focus on the impact of the NASA Landsat satellite mapping program on shaping the level and distribution of new discoveries between firms in the gold exploration industry. I find that idiosyncratic gaps in mapping coverage (from technical failures and cloud-cover in satellite imagery) had important implications for gold exploration— firms were almost twice as likely to report the discovery of new deposits once regions were successfully mapped and the mapping program disproportionately supported discoveries from smaller, entrepreneurial firms, especially in regions with high quality local institutions. These findings point to the important but underexamined role of mapping as an economic activity in shaping industry performance and entrepreneurship.

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

Fundamental and basic knowledge about the physical world has led to new discoveries and massive increases in human prosperity since the middle ages (Romer, 1990). Economic history indicates that an important channel through which basic knowledge about the physical world could have enabled discovery is through novel maps of poorly understood geographies (Whitfield, 1998). For example, the *Itenerario*, a compendium of maps published in 1596 by the merchant Jan Huyghen Van Linschoten, contained 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" (Davids, 1986). Soon after this new map was published, the Dutch and British East India companies were established, many new territories and trading partners were discovered and the Portuguese monopoly over the trade and colonization in south-asia was ended (Saldanha, 2011; Jefferson, 2013). This anecdote begs the question: how does the arrival of basic knowledge through the publication of new maps causally affect the discovery of new opportunities and entrepreneurship in the private sector? This paper investigates this question in a modern context— the role of the NASA *Landsat* satellite mapping program in shaping the discovery of new deposits in the gold exploration industry.

Despite its status as one of the oldest forms of basic knowledge, the possible role of mapping information in shaping the geography of private discovery and entrepreneurship has resisted formal investigation. This is surprising because, unlike Borges' fantasy from the quote above, it is a cartographic truism that "there is no such thing as a complete map" (Harley, 1989). In practice, even after a region has been mapped, it is quite common for many territories to have been ignored or poorly understood (Monmonier, 1991). While this variation in the availability of basic geographic knowledge across regions is quite prevalent, whether and how it affects the level and distribution of performance between larger and smaller firms is hard to know ex-ante. On one hand, the private sector has invested significantly for hundreds of years in mapping different regions around the world 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 (Budish et al., 2013; Nelson, 1959), then the lack of mapping

information in some regions relative to others could significantly boost discovery in regions that benefit from new knowledge. While both arguments have theoretical validity, without empirical investigation it is difficult to evaluate whether public investments in new maps stimulate new discoveries. Further, in addition to their affecting new discoveries at the industry-level, theoretically it is plausible that mapping information could disproportionately affect larger or smaller market participants (Arora and Cohen, 2015). On one hand, public investments in basic knowledge could help larger firms reinforce market power– for example, through their capabilities to better absorb external knowledge (Cohen and Levinthal, 1990). On the other hand, public mapping information could boost entrepreneurship by disproportionately helping smaller firms, for example, by reducing the uncertainty of early exploration (Kerr et al., 2014). While strong theoretical arguments exist for both effects, without empirical evidence it is difficult to shed light on this second question of how new maps affect the distribution of industry performance between larger and smaller firms.

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 public investments in basic knowledge on both industry performance and entrepreneurship. Specifically, I study the Landsat satellite mapping project and its role in shaping the discovery of new deposits in the gold exploration industry. Landsat provided the first images of Earth from space, and while the program was designed for its agricultural (and not geological) applications, maps from the program provided information that was relevant to guide early-stage gold exploration (Rowan et al., 1977). Further, the Landsat program is a particularly appropriate setting because it represents a natural experiment with plausibly exogenous allocation of mapping information to some regions and not others. Specifically, 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 different regions. Of the 9493 "blocks" (regions of 100 sq. mile each) which are needed for full coverage of the earth, some blocks received satellite maps early in the program, while others were mapped at significantly later points in time over the next decade. Further, quantitative assessments and qualitative interviews indicate that significant differences in the timing of the mapping effort were unintentional on the part of the program administrators, due to reasons like technical failures in satellite operation and cloud-cover in imagery.

I utilize this variation to estimate the impact of the Landsat program on the gold exploration industry. The economic importance of the gold exploration sector (with approximately $5 billion was spent on gold exploration in 2010 alone (Schodde, 2011)) and availability of detailed data on discoveries makes it an especially attractive industry for this study. Conceptually however, the findings from this study could generalize to exploration for other natural resources like oil and gas, copper, and uranium. To causally isolate the impact of Landsat on gold discovery, I exploit the quasi-random variation in the timing of the mapping effort using a differences-in-differences

framework. Importantly, this specification flexibly controls for differences in "prospectivity" (the "true" probability of finding resources) between different regions through block fixed-effects, and for secular changes in the gold exploration market over time through year fixed effects. The key assumption required for this specification— similar changes in discovery over time— holds in the data. Further, an instrumental-variables (IV) specification that uses the cloudiness of different regions as an instrument for the timing of the mapping effort provides a robustness check for the main specification. The main dependent variable in these specifications, an indicator variable for new gold discoveries at the block-year level, is obtained from a unique, hand-collected database of major discoveries by exploration firms between 1950 and 1990.

The results suggest that, despite strong private incentives for mapping, the public Landsat mapping effort 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. These differences imply meaningful benefits of the mapping effort in dollar terms— 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 block over a fifteen year time period. For a country the size of the United States, this translates to additional gold reserves worth about $10 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.)

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. Specifically, I test whether the mapping program disproportionately benefited "juniors", smaller and entrepreneurial firms in the exploration industry as compared to "seniors," larger and more established players, to understand whether public information helps boost the performance of smaller firms or whether it reinforces the position of dominant firms. I find that the smaller firms share an increased proportion of the new discoveries attributed to the Landsat program as compared to before the launch of Landsat. Specifically, 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. Put differently, junior-led discoveries increased by a factor of 5.8, while the corresponding rise for seniors was only about 1.7, indicating that smaller firms benefited more than three times as much as incumbents from new mapping information. These results suggest that mapping information both raises the overall level of industry performance and disproportionately encourages the performances of smaller firms. However, the hypothesis that the Landsat program completely displaces incumbent senior firms does not seem to find support in the data.

Finally, in order to examine further the conditions under which new maps could encourage entrepreneurship, I rely on the Fraser Institute Survey of Mining Companies, which records data on the quality of institutions around the world relevant to the exploration sector. I find that when local policies support the mining industry through institutions like strong property rights and clear legal and labor market regulations, the impact of the mapping program is almost three times as strong, suggesting complementarities between investments in basic knowledge and institutions. Along these lines, when regions have a poor overall level of institutional quality through factors like uncertainty in regulations and political instability, the mapping program promotes discovery by seniors but does not help junior-led discovery. In other words, mapping information does not substitute for a lack of institutional capacity in regions around the world, but instead reinforces the impacts of institutional differences on entrepreneurship.

This paper contributes to the literature on the role of public investments in knowledge goods on encouraging the performance of firms in the private sector. While this literature is fairly extensive (see Czarnitzki and Lopes-Bento (2013) and David et al. (2000) for an overview), this study joins three recent papers in evaluating this question using exogenous changes in the level of public investments on private patenting (Moretti et al., 2014; Azoulay et al., 2015) and entrepreneurship (Howell, 2014). The present study adds to this literature by highlighting a novel channel through which openly-available knowledge could matter for industry (investments in mapping goods), and by focusing on a direct measure of firm performance (discovery), rather than intermediate measures of performance such as patenting. This is also the first paper, to my knowledge, that finds that public information could differentially affect the performance of larger and smaller firms.

This paper also builds on a secondary literature in cartography and management that focuses on the importance of the activity of mapping. The literature in cartography has long identified the importance of the details of the map-making process for the quality of the resulting map (Guo, 2011; Dodge et al., 2011; Monmonier, 1991). While this literature assumes that mapping quality has large implications (Crampton and Elden, 2007; Harley, 1988), this paper provides direct and causal evidence for the role of mapping on important economic outcomes. Similar to the cartographic literature, research in the area of managerial cognition (Kaplan, 2008; Tripsas and Gavetti, 2000) and "sensemaking" (Weick, 1995), has theoretically studied "mental maps" (Puranam and Swamy, 2010) and has paid attention to the role of representation (Gavetti et al., 2004; Gavetti and Levinthal, 2000) and cognitive mapping on firm performance. There is also some work in political science on the role of mental mapping on the political decisions of policy-makers (Axelrod, 1976) and work in economics and sociology on the role of representation in knowledge production (Latour and Woolgar, 2013; Cowan et al., 2000). I contribute to this work by providing an empirical framework to assess theoretical propositions in the literature by focusing on map-

making in a physical and literal sense.

Finally, this paper also contributes to the literature on the role of institutions in determining firm performance and entrepreneurship (Tolbert et al., 2011). A vast literature in economics, sociology, and management has studied the role of formal and informal institutions on entrepreneurship and performance (Kerr and Nanda, 2009; Khessina and Carroll, 2008; Sorenson and Audia, 2000). I contribute to this work by uncovering a novel channel through which institutions could affect firm outcomes, namely influencing the ability of firms to benefit from public investments in knowledge.

2 Conceptual Framework and Literature

To understand how exactly mapping could affect discovery consider the following conceptual framework. Consider two regions, A and B, both with similar expected prospectivity in terms of the likelihood of discovering new gold deposits. Firms invest to generate private maps of these regions to guide exploration, and perhaps also explore and make new discoveries. Now consider that a public mapping program such as Landsat provides freely available mapping information for region A and not for region B. How will the difference in the availability of information affect exploration firms interested in finding new deposits in these two regions? Specifically, how it will affect the likelihood of making a discovery in region A as compared to B, and will mapping affect larger and smaller firms differentially?

In this conceptual framework, I will focus on two factors through which the mapping program could affect industry outcomes: cost of private mapping efforts and the effectiveness of new maps in resolving uncertainties in the exploration process. First, if the public sector fails to provide mapping information in some regions, firms might not be affected because they have already made investments in private maps, and already possess detailed knowledge of the underlying potential of a given region. The ability of firms to make such investments depends heavily on the costs of producing private mapping information. If these costs are low then we should expect that poor public mapping should have little impact on private discovery because firms would have already made such investments. However, if such costs are significant or alternative mapping methods are not available, then we should expect that variation in public mapping should negatively impact the likelihood of making discoveries in unmapped regions. Further, if the costs are low for larger firms but are prohibitive for smaller firms, then we might expect the public information to enable discoveries by smaller firms, but not affect larger firms substantially (Cohen and Klepper, 1996).

Second, the overall impact of missing mapping information depends on the relative informativeness of the public map. If public maps help to clarify uncertainty, i.e. they help more precisely identify

promising regions for discovery and rule out unattractive regions, then we should expect that public mapping increases the likelihood of discovery in mapped regions. However, if Landsat provides information that does not provide previously unknown information, perhaps because it was not designed for the mining industry, then we should expect its impact to be fairly small. Similar to the argument for mapping costs, the informativeness of the public map could also be different for larger and smaller firms. On one hand, larger firms might have better ability to understand and interpret Landsat maps because of their pre-existing connections to external knowledge (Cohen and Levinthal, 1990; Sosa, 2009) while smaller firms may find it to be less useful. On the other hand, smaller firms might be more likely to be on the technological frontier (Schumpeter, 1909) or might be more likely to value the new technology (Benner and Tripsas, 2012; Christensen, 1993) while larger firms might be unable to benefit given organizational rigidities (Sosa, 2012). These differences in the perceived informativeness of the new map might create performance differences between larger and smaller firms. Therefore, the overall impact of the mapping program depends on the relative importance of these two channels– the costs of private mapping efforts and the relative informativeness of the public map to larger and smaller firms.

Both of these margins are relevant in the case of Landsat. As far as private investments in mapping information is concerned, there exist numerous other sources for exploration firms like aerial imagery, ground surveying and soil and stream sampling. Often, such alternate mapping information is available for free or at reduced costs from industry agencies and associations. Further, while the NASA Landsat mission was the only provider of satellite imagery, there were commercial providers of such imagery that emerged in the mid 1980s. However, a privately-funded mapping mission or extensive aerial mapping program could have been quite expensive. Further these alternate private mapping efforts were often quite expensive, and while commonly employed by larger firms, were often not employed by smaller firms. Smaller firms depended on publicly available information, or they would rely on the founders' private knowledge gained from their social networks or from previous work experience.

Second, it was unclear how effective Landsat imagery was in resolving exploration uncertainties. On one hand, space mapping technology like Landsat provided quite coarse and low-resolution imagery that was unsuitable for small-scale mapping. However, satellite imagery did offer maps of wide swathes of the surface of the earth which allowed for the identification of previously unknown features that could have improved geological exploration models, and reduced uncertainty. This value of this new information could have varied by larger and smaller firms. On one hand, larger firms usually have large exploration divisions that are connected with NASA and academic researchers pioneering the use of this new information, and could have learned about the Landsat earlier and more effectively as compared to smaller firms. On the other hand, the organizational

structure of exploration at larger firms often relied on on-the-ground prospectors, and these investments could have created disincentives to invest in evaluating the value of maps from the Landsat program.

This conceptual framework clarifies mechanisms through which public mapping programs could influence exploration by firms, and affect discovery. While there is no direct literature on the role of mapping information on economic outcomes that I am aware of, the framework presented above is related closely to literature on the role of openly provided basic knowledge on innovation and performance at the firm-level (Hall and Van Reenen, 2000). At the core of this literature is the idea that the private sector might not have sufficient incentives to invest in basic knowledge (Arrow, 1962), despite significant productivity benefits from such information. This market failure could stem from two different channels (Nelson, 1959). First, investments in basic knowledge are often speculative and it is difficult to predict ex-ante whether and how such information will prove useful to the investing firm. Second, basic knowledge often tends to have a wide applicability across industries, and it is difficult for the investing firms to internalize externalities from knowledge spillovers to other firms or industries from basic knowledge. Combined, these two channels imply that, theoretically, there might be significant underprovision of basic knowledge.

Despite the prevalence of this argument, empirical evidence for the predictions of this theory have been mixed and inconclusive (David et al., 2000), largely due to empirical challenges. In order to test the predictions of the theory, it is necessary to introduce quasi-random variation in the amount of basic knowledge and estimate the impact of such variation on firm performance and innovation. While a number of different papers have tried to estimate regressions in this spirit (see Czarnitzki and Lopes-Bento (2013) or Hall and Van Reenen (2000) for a review), empirical challenges have complicated interpretation.

Two recent papers identify a more precise relationship between investments in basic knowledge and firm innovation using a more robust quasi-experimental research design. Azoulay et al. (2015) introduces variation in investments across different research areas using rules governing the National Institutes of Health (NIH) peer review process. They study the impact of funding variation on patenting by private sector firms and find that a $10 million boost in NIH funding leads to an increase of 2.3 patents by the private sector. Another related paper, Moretti et al. (2014) estimates the impact of exogenous changes in defense R&D expenditures on private sector R&D. They find a positive relationship between public support for R&D and industry productivity, implying that the provision of basic knowledge might have an important role to play in boosting industry performance.

My paper builds on this literature and tests whether investments in basic knowledge are able to boost private discovery for larger and smaller firms in a novel context— the gold exploration indus-

try. This setting allows me to contribute to this literature by examining the impact of basic knowledge on a concrete measure of performance (the discovery of new deposits) in a quasi-experimental setup. Further, I'm also able to quantify how the impact of the new investments could vary by larger and smaller firms who could benefit from public knowledge.

3 Empirical Setting

3.1 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 covers an area of about 185km × 185km, and 9493 satellite images are required to cover all of Earth's land-masses (not including Antarctica and Greenland). 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.

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. Landsat imagery was relayed to the Earth Resources Observation and Science (EROS) center in Sioux Falls, South Dakota which was established to collect and distribute these data to follow-on investigators. The EROS data center distributed data 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. 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). The EROS repository helps me track information about the satellite images directly, including the location of blocks, when they were mapped, and the quality of the mapping effort.[1]

---

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

3.2   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 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 enabled academic 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. Whether this information was previously unknown to exploration firms, and whether they found it useful remains an empirical question.

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, firms often sourced aerial imagery from mapping surveys conducted from airplanes. In fact, firms routinely collected aerial mapping information, but systematic and country-wide aerial mapping programs were rare (Spurr, 1954). Further, it was also possible to replicate Landsat information by launching a new, satellite-based mapping program in the private sector. 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.

4   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, I am also interested in collecting measures of the quality of different local institutional policies across different regions. This section describes the study's data collection process in further detail.

4.1 Data

A. Landsat Coverage Data: I construct data on Landsat coverage from the USGS EROS data center's sensor metadata files.[2] These data provide a list of all images collected by the Landsat sensors, including the location of the image, 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.

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. While this database is unlikely to have 100% coverage, estimates suggest that about 93–99% of all valuable discoveries are included. See data appendix for more details about this data source. Using micro data on all available discoveries, I first match them to a specific block-year using geographic coordinates. Having performed this matching, I then aggregate all discoveries within a given block-year and conduct my analysis at this level. The main outcome variable is $Any\ Discovery_{it}$ which is an indicator variable for whether a discovery was made in a given block-year. In total, 414 unique blocks have reported a total of 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.

---

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

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 compile a list of all publications related to gold exploration from Scopus. 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 article, typically the region of the field site of the study. Using the latitude and longitude of each entity and the date of the publication, I link the observation to a block-year observation in my dataset. Using these data, I 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 level of scientific research about a given block in a given year.

Second, I use the "Global Earthquake Hazard Frequency and Distribution" database (Dilley et al., 2005; Center for Hazards and Risk Research - CHRR - Columbia University and Center for International Earth Science Information Network - 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). 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.

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). I match these data to create a measure of average cloud cover percent corresponding to each Landsat block. This measure is employed in my instrumental variables specifications.

D. Mining Institutions Survey: In order to explore the impact of local institutions in influencing the impact of Landsat on the mining industry, I rely on the "Survey of Mining Companies" conducted by the Fraser Institute.(McCahon and Fredricksen, 2014) While the survey has been conducted annually since 1997, I use the 2014 edition[3], which contains information on over 122 dif-

---

[3]A key assumption is that local institutional conditions do not change significantly in response to variation in the Landsat effort. Despite this concern, I employ the 2014 edition of the survey because it has the maximum coverage across regions, and there were no similar surveys conducted before the Landsat program was launched.

ferent jurisdictions around the world – including provinces in major mining countries like Canada, Australia, USA, etc. The survey contacted about 4200 managers and executives in the area of mining exploration and received 485 responses (response rate of 11.5%) on which the survey is based. The firms in the survey reported an exploration spend of about \$2.5 billion in 2014[4], and represented nearly all significant organizations in the exploration industry.

The survey was designed to capture the opinions of managers and executives about the level of investment barriers in jurisdictions with which their companies were familiar. They were asked 15 questions about the quality of different local institutions that affect mining investment and were asked to rate each one on a scale of 1 through 5, in terms of whether they "encourage exploration investment" on one end or whether local institutions were a "strong deterrent to exploration investment" on the other. This survey includes a number of questions about factors like uncertainty about the legal system, taxation regime, land claims and property rights at the regional level. The responses are used to rank each of the 122 jurisdictions on a "policy perception index" that I use as a measure of the quality of local institutions as it relates to mining activity. For each Landsat block, I use this rank measure to operationalize a regional variable that captures the quality of local institutions and their ability to encourage mining investment. A complete list of the 122 regions in the survey, along with their individual ranking is provided in Appendix Table C.6.

E. Summary Statistics: Table 1 provides a list of key variables used in the quantitative analysis 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, 0.19, can be interpreted as the percentage probability that a discovery is reported in a block-year.[5] *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

---

[4]The total industry expenditure on gold exploration was about \$4.5 billion dollars in 2014(Carlson, 2014).

[5]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.

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%. Please see the data appendix for more details on the sources of data and the data construction process.

4.2 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.

A. Differences-in-Differences Specification: In order to implement the differences-in-differences specification, I first establish (in the next section) large variations in the amount of Landsat imagery that was available for distribution in different regions of the world. A simple comparison of the trend of gold discoveries in regions with intensive 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 intensively mapped regions 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 using quite limited assumptions.

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. 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.

4.3   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 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 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 is 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 a greater intensity of Landsat coverage as compared to other blocks is represented in Figure 1. The number of images here is simply a proxy for the quality of mapping information provided at the block-level. Above-median blocks are all blocks that receive more than twenty images by 1983, while below-median blocks receive twenty images or less. As the figure illustrates, discoveries in above-median and below-median blocks had a 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 above-median blocks show a much faster rate of growth as compared to below-median blocks. The difference in gold discoveries between these two groups appears to vary by a factor of three by 1990. This analysis provides some preliminary evidence to suggest that Landsat mapping had a large impact on gold discovery. However, while there do not seem to be any trend differences in trends between above-median and below-median 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 above-median and below-median blocks. The data indicate that above-median blocks are indeed more likely to report new discoveries of gold before 1972, have higher prospectivity scores, and are more likely to have gold-mining related publications. While in theory these level differences could be controlled for using block fixed-effects, the baseline DD specification that compares blocks mapped early to those mapped at a later point in time could potentially provide more reliable estimates of the impact of Landsat on discovery. To see this, consider Figure 3, which compares blocks mapped in the first two years of the Landsat mission (before 1974), with blocks mapped later (after 1974 or never). These charts not only indicate no changes in trends between these two groups of blocks in terms of discoveries, publications or the probability of earthquakes, but also indicate no level differences. This evidence provides us with confidence that the main difference-in-difference specification is providing estimates from reasonably comparable groups.

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.[6] This analysis provides some preliminary

---

[6]There seems to be some difference in the number of publications. However, the difference is in the opposite

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.

5   Results

5.1   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 land-masses 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. Further, the distribution of the outcome variable (unreported) suggests that most block-years report either no discoveries or at most one discovery. There is a small number of cases when block-years report more than one discovery. Motivated by this distribution,

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.

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) models instead of count data models (like poisson quasi maximum likelihood models) which are quite popular in the innovation literature. 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.

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. The baseline results therefore confirm the main hypothesis that the Landsat mapping effort had a significant impact on industry performance.

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.[7]

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 two 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 – a reasonable estimate of discovery timelines in the gold exploration industry (Branch, 2009).

C. Instrumental Variables evidence: Now, I turn to analyzing the impact of Landsat coverage

---

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

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. In this section, I use 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. The figure 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 quality of the best available image at the block level. This scatterplot also 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 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. The figure shows that there is a slightly positive or flat relationship between the two variables. 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. These plots provide strong support for the exclusion restriction of the cloud cover instrument.

Table 4 provide 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. 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 1.051— 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 6.5x 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, or perhaps because the instrument is inappropriate and fails the exclusion restriction.

5.2 Did Landsat Democratize Discovery? Comparing Different Types of Firms

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 now turn to analyzing the role of satellite mapping on incumbents and startups in the industry.

In this section, I will 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. Specifically, 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 0.04, 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.

Figure 6 plots these gains in percent terms. As this figure illustrates, 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. This evidence helps guide theory on the role of mapping on the performance of startups and incumbents. In this case, it seems like new mapping investments, disproportionately helped the performance of smaller firms in the gold exploration industry, and did not reinforce the market power of seniors, suggesting that investments in new maps could not only raise overall performance at the industry level, but that they could also help encourage entrepreneurship.

5.3 The Role of Local Institutions

Finally, I investigate the conditions under which startups can translate opportunities from new maps into improved performance in the exploration industry. For this analysis, I rely on 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. The responses are used to rank regions based on a composite score that captures the performance of different regions in terms of institutional quality.

I employ this rank measure and classify regions into three equally sized categories— high-quality, medium-quality and low-quality— based on their rank in the Fraser Institute survey. Having classified regions in this way, I estimate the following specification:

$$Y_{it} = \alpha + \beta_1 \times Post_{it} \times 1(High_i) + \beta_2 \times Post_{it} \times 1(Medium_i) + \beta_3 \times Post_{it} \times 1(Low_i) + \gamma_i + \delta_t + \epsilon_{it}$$

Similar to the baseline specification, γ_i and δ_t represent block and time effects respectively for block i and year t and $Post_{it}$ is an indicator variable denoting whether a region has been mapped with a low-cloud image. $1(High_i)$, $1(Medium_i)$ and $1(Low_i)$ are indicator variables that are set to one if the region is in the first, second or third tercile of the policy rank distribution according to the Fraser Institute survey.

The estimates from this regression are presented in Table 6, and the corresponding elasticities are indicated in Figure 7. These results suggest that the overall positive impact of startups on capturing new opportunities from the Landsat program are increasing in the quality of local institutions – i.e. in the top tercile of regions where local institutions support mining operations, startups benefit at a rate of 15x compared to incumbents who benefit at 6.6x. However, in the medium and low end of the quality distribution, these differences are reduced. In the bottom quartile, incumbents are still benefiting from government investments (at the rate of 0.6x), while startups seem to be negatively impacted by new information. The significantly negative impact of new information on junior-led discoveries in regions with low institutional quality is quite puzzling. One likely explanation is the

substitution of junior search from regions of low institutional quality into regions with high quality of institutions, which could explain this finding. However, this hypothesis is tentative and needs further examination.

These results suggest that while Landsat helped the performance of startup firms in the gold exploration industry, the role of local institutions was large complementary to this process. In other words, to encourage entrepreneurship, it is not sufficient to only provide investments in new mapping information— the support of local institutions has an important complementary role to play.

6   Discussion

This paper studies the role of the open provision of mapping information and its impact on the performance of firms in the private sector. It highlights the process of map-making as a commonly employed, but rarely studied channel through which geographic variation in performance and entrepreneurship can be explained. The paper relates to the literature on the open and public provision of knowledge goods, specifically the extensive literature examining different mechanisms for public sector involvement in addressing deficiencies in innovative activity by firms, including studies of R&D tax credits, subsidies, intellectual property, grants, and prizes. While some of these measures have been shown to be quite effective in stimulating knowledge production and firm performance, the evidence for the effectiveness of these measures is quite mixed. In this paper, I suggest an alternate and under-studied mode through which the public sector could incentivize innovation, i.e. the production of "basic" mapping datasets that are openly available to end users. The main contribution of the paper is to highlight the specific details of mapping programs could have large impacts for industry performance, and could therefore be strategically used to shape the performance of firms.

Specifically, I study Landsat, a NASA satellite mapping technology and its impact on the gold exploration industry. In this industry, Landsat maps provided early-stage exploration guidance that could be used by firms to help the discovery of new mineral resources. I exploit quasi-random variation in the timing and coverage of the Landsat mapping effort to measure the impact of the arrival of new maps on the discovery of new deposits in mapped and unmapped regions. I find that Landsat doubled the likelihood of making new discoveries in mapped regions, as compared to unmapped regions. Further, the data indicate that startups were significantly more likely to benefit from the new maps as compared to incumbents. Finally, gains in startup performance were heavily concentrated in regions with higher quality local institutions that support mining, suggesting that

public knowledge goods must be complemented by effective local institutions to encourage startup performance.

This paper makes a few contributions to our understanding of the effect of public investments in knowledge goods on private performance. First, I document the performance implications of the provision of public knowledge goods for industry in a quasi-experimental framework. This adds to recent evidence (Howell, 2014; Azoulay et al., 2015) on the causal role of public funding on private performance. This study adds to this literature in two important ways. First, I am able to estimate the impact of public financing of knowledge on market structure. It seems that government investments in knowledge do not just increase performance uniformly, but could also disproportionately favor the performance of smaller firms in the industry. Second, I am also able to link the provision of new information to commercial outcomes of economic interest to policymakers (the discovery of a new deposit). This paper is also related to the literature on the role of open access knowledge goods on innovation (Williams, 2013; Furman and Stern, 2011). This work has demonstrated quite convincingly that institutional features of the information environment could have important implications for the diffusion of new knowledge. I add to this literature by pointing out the ways in which different types of firms could be impacted differently by the provision of new information.

This paper builds upon insights from a wide variety of different social sciences that have touched upon the topic of mapping and cartography. This includes the literature on geographic information systems (GIS), on the design and representation of space using maps (for example, (Monmonier, 1991)), and the literature in management on cognition and mental mapping (Gavetti et al., 2004). There is also a broader literature on tacit knowledge, knowledge representation and the role of the representation of information that this paper touches upon (for example, (Axelrod, 1976) in political science), that this paper builds on. Finally, this paper also contributes to existing literature on the role of institutions on entrepreneurship (Tolbert et al., 2011). While institutions could foster entrepreneurship by reducing the costs or risks of entry, this paper highlights another channel through which institutions could boost entrepreneurship. Specifically, institutions could help startups identify and act upon opportunities from public datasets, and therefore support startup performance. The importance of local institutions is therefore magnified during periods of technical change in industries.

A few different policy recommendations follow from these findings. My results suggest that investments in public mapping goods can have significant and positive impacts for industry. Second, these findings highlight the role of local institutions in the context of fast-changing and high-technology industries. In these cases, strengthening local institutions will not only increase overall perfor-

mance, but will also help boost entrepreneurship in innovative markets. Finally, these results also point to the international nature of public information policy. Landsat seems to have caused significant changes in discoveries, and possibly regional economic activity, across different parts of the globe, even though it was a program that was designed and developed for the US. This points to the increasingly international nature of information policy, suggesting a role for information policy on issues related to globalization and international development.

Finally, the findings from this paper could generalize to settings where maps don't just represent physical space and geography, but could also be used to represent other real-world objects. Examples include the Human Genome Map (Williams, 2013), or projects like the mapping of space or the "brain mapping" project. In fact, in strategy, an open question is the extent to which managers are able to map the environment about them, and the implications of different forms of representations of similar spaces for performance. The empirical framework provided in this paper, could apply to these general settings as well. Whether by shaping entrepreneurs' or managers' mental maps of the environment, their decision-making and firm performance could be positively or negatively influenced remain open, but exciting questions in this area of research.

References

Arora, A. and W. M. Cohen (2015, July). Public support for technical advance: the role of firm size. *Industrial and Corporate Change*, dtv028.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pp. 609–626. Princeton University Press.

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,.

Axelrod, R. (1976). Structure of decision: the cognitive maps of political elites.

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.

Benner, M. J. and M. Tripsas (2012). The influence of prior industry affiliation on framing in nascent industries: the evolution of digital cameras. *Strategic Management Journal 33*(3), 277–302.

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

Budish, E. B., B. N. Roin, and H. Williams (2013). Do Fixed Patent Terms Distort Innovation?: Evidence from Cancer Clinical Trials. *Evidence from Cancer Clinical Trials (September 5, 2013). Chicago Booth Research Paper* (13-79).

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

Center for Hazards and Risk Research - CHRR - Columbia University and Center for International Earth Science Information Network - CIESIN - Columbia University (2005). Global Earthquake Hazard Frequency and Distribution.

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

Christensen, C. M. (1993). The rigid disk drive industry: a history of commercial and technological turbulence. *Business history review 67*(4), 531–588.

Cohen, W. M. and S. Klepper (1996). Firm size and the nature of innovation within industries: the case of process and product R&D. *The review of Economics and Statistics*, 232–243.

Cohen, W. M. and D. A. Levinthal (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly*, 128–152.

Cowan, R., P. A. David, and D. Foray (2000). The explicit economics of knowledge codification and tacitness. *Industrial and corporate change 9*(2), 211–253.

Crampton, J. W. and S. Elden (2007). *Space, knowledge and power: Foucault and geography.* Ashgate Publishing, Ltd.

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, 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.

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.

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.

Dodge, M., R. Kitchin, and C. Perkins (2011). *Rethinking maps: new frontiers in cartographic theory.* Routledge.

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.

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.

Gavetti, G. and D. Levinthal (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative science quarterly 45*(1), 113–137.

Gavetti, G., D. Levinthal, and J. W. Rivkin (2004). *Strategy-making in novel and complex worlds: The power of analogy.* Division of Research, Harvard Business School.

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*.

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.

Guo, Z. (2011). Mind the map! The impact of transit maps on path choice in public transit. *Transportation Research Part A: Policy and Practice 45*(7), 625–639.

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.

Harley, J. B. (1988). Maps, knowledge, and power. *Geographic thought: a praxis perspective*, 129–148.

Harley, J. B. (1989). Deconstructing the map. *Cartographica: The international journal for geographic information and geovisualization 26*(2), 1–20.

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.

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).

Kaplan, S. (2008). Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. *Academy of Management Journal 51*(4), 672–695.

Kerr, W. R. and R. Nanda (2009). Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship. *Journal of Financial Economics 94*(1), 124–149.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014). Entrepreneurship as experimentation. Technical report, National Bureau of Economic Research.

Khessina, O. M. and G. R. Carroll (2008). Product demography of de novo and de alio firms in the optical disk drive industry, 1983-1999. *Organization Science 19*(1), 25–38.

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,.

Latour, B. and S. Woolgar (2013). *Laboratory life: The construction of scientific facts*. Princeton University Press.

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

Monmonier, M. (1991). How to Lie with Maps.

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.

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

Puranam, P. and M. Swamy (2010). Expeditions without Maps: Why Faulty Initial Representations May Be Useful in Join Discovery Problems. *Available at SSRN 1153142*.

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.

Saldanha, A. (2011, January). The Itineraries of Geography: Jan Huygen van Linschoten's Itinerario and Dutch Expeditions to the Indian Ocean, 15941602. *Annals of the Association of American Geographers 101*(1), 149–177.

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

Schumpeter, J. (1909). On the concept of social value. *The quarterly journal of economics*, 213–232.

Sorenson, O. and P. G. Audia (2000). The social structure of entrepreneurial activity: geographic concentration of footwear production in the United States, 194019891. *american Journal of Sociology 106*(2), 424–462.

Sosa, M. L. (2009). Application-specific R&D capabilities and the advantage of incumbents: Evidence from the anticancer drug market. *Management Science 55*(8), 1409–1422.

Sosa, M. L. (2012). Organizational Structure, Real Options and the Advantage of De Novo Firms. *Organization Science*.

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

Tolbert, P. S., R. J. David, and W. D. Sine (2011). Studying choice and change: The intersection of institutional theory and entrepreneurship research. *Organization Science 22*(5), 1332–1344.

Tripsas, M. and G. Gavetti (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic management journal 21*(10-11), 1147–1161.

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.

Weick, K. E. (1995). *Sensemaking in organizations*, Volume 3. Sage.

Whitfield, P. (1998). *New found lands: maps in the history of exploration*. Psychology Press.

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.

7 Figures and Tables

Figure 1. **Mean probability of gold discoveries for blocks with above-median and below-median levels of Landsat coverage**
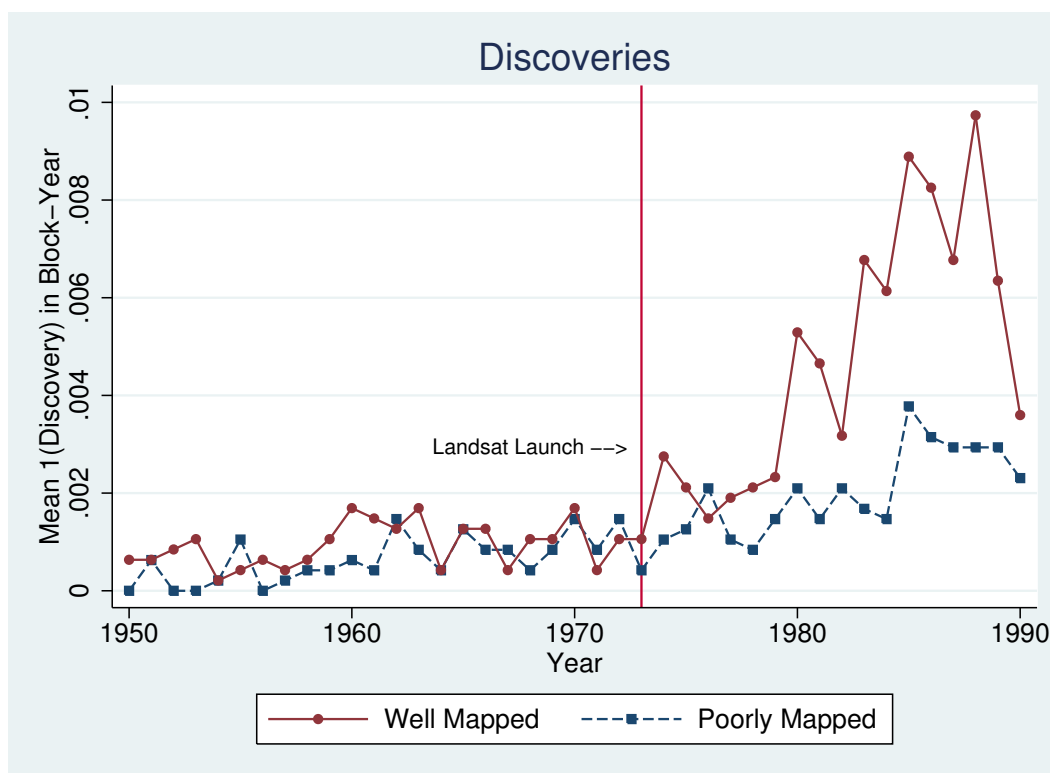
Note: This figure plots the history of global gold discovery over time, using block-year level discovery data. Each block is classified either as a "Well Mapped" coverage block (in red circles), or "Poorly Mapped" block (in blue squares), depending on whether it received above or below the median number of images from the Landsat program (between 1972 and 1983). 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 the variation in Landsat coverage**
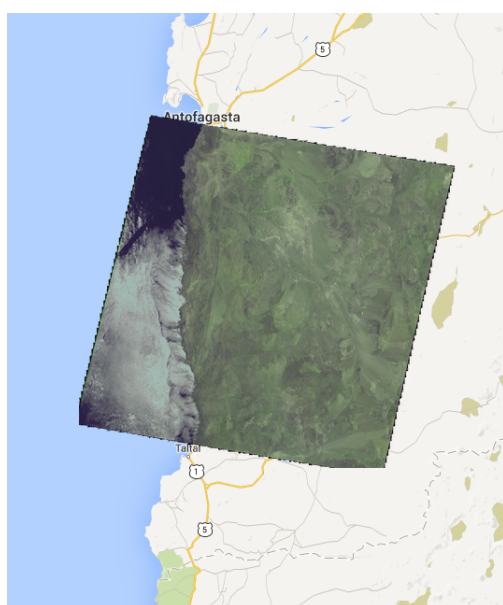
(1) Block 25177, Chile
Cloud free imagery available
Amax Gold Discovery reported in 1980
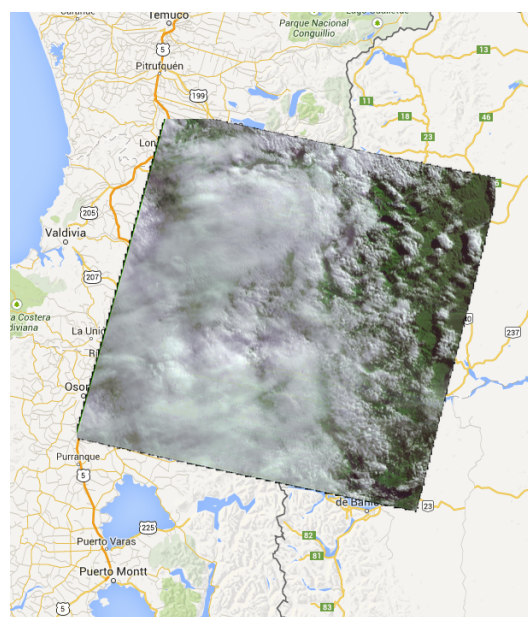
(2) Block 24988, Chile
No cloud free imagery available by 1983
No discovery reported to date

Note: This figure provides an example that illustrates the research design and data used in this paper 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 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. As depicted, the best available image has considerable cloud-cover, and no low-cloud image was available by 1983 for this block and no discovery has also been reported in this region to this date.

Figure 3. **Comparing blocks mapped early and late before Landsat launch**

Panel A. Average Annual 1(Discovery)
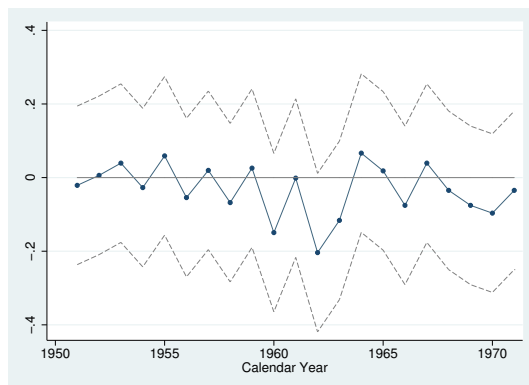
Panel B. Average Annual Publications
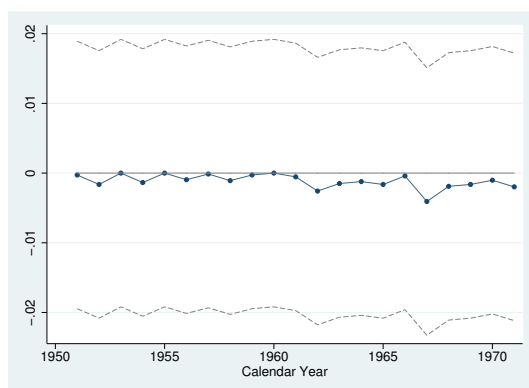
Panel C. Average Annual Earthquakes
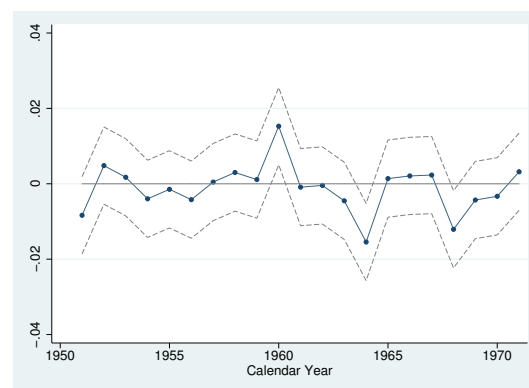
Note: This figure explores pre-treatment differences in early and late-mapped blocks. For each of three figures, difference in means of outcome variable is calculated between blocks mapped early (mapped before 1974) with blocks mapped late (on or after 1974) on a yearly basis. The outcome variables are average of indicator variable for discovery in block-year (Panel A), average gold-related publications in mining journals (Panel B) and average number of earthquakes (Panel C).

Figure 4. **Time Varying Estimates of the Impact of Landsat Intensity on Gold Discovery**
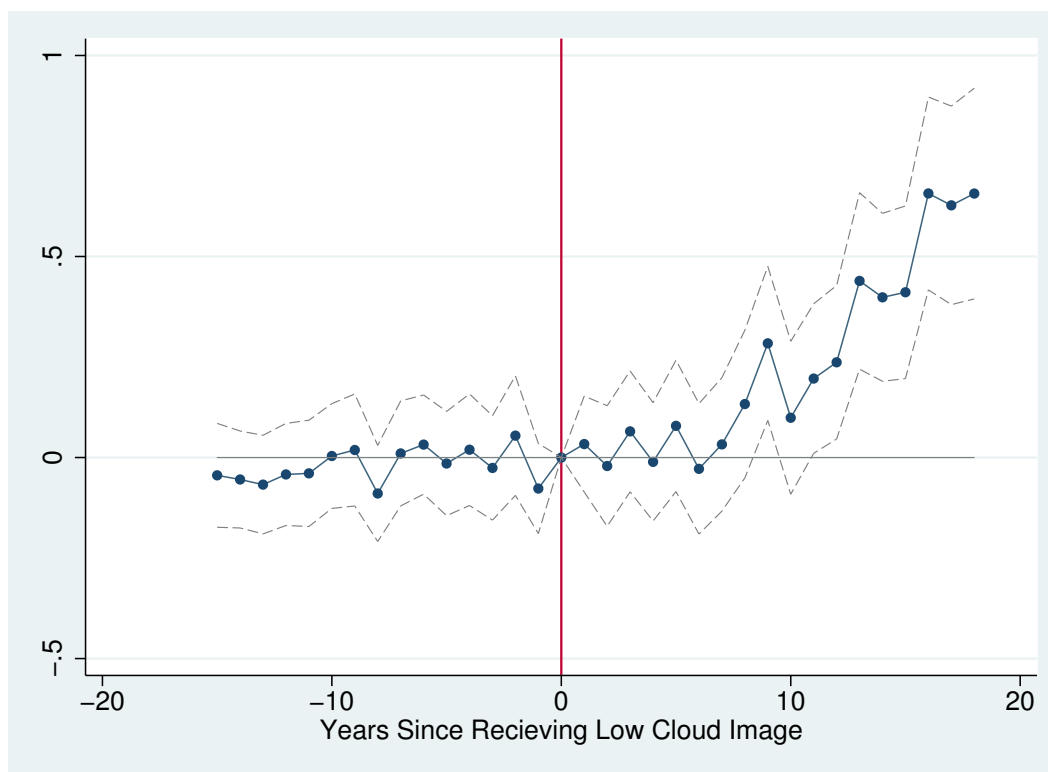
Note: This figure plots estimates (and 95 percent confidence intervals) of β_t from the event study specification specified below. On the x axis is calendar year. This figure is based on block-year observations, the coefficients are estimates from OLS models, the sample includes all block-year discoveries 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. **Testing the Cloud Cover Instrument**

Panel A: First Stage: cloud cover and Image Timing
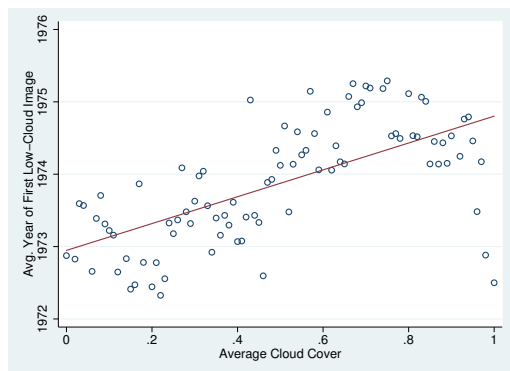
(1) Avg. Year of First Low-Cloud Image
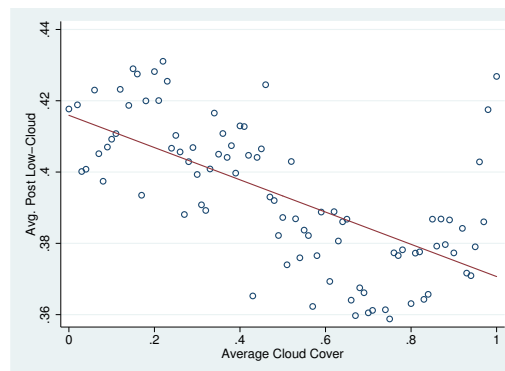
(2) Avg. Post Low-Cloud Indicator

Panel B: Exclusion Restriction: cloud cover and Correlates of Gold Discovery
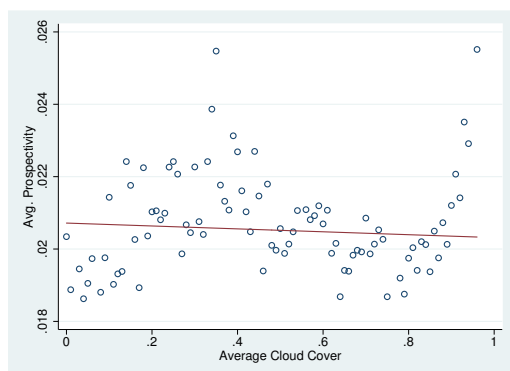
(1) Avg. Predicted Prospectivity
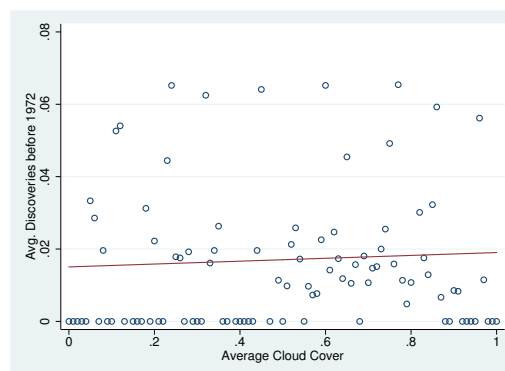
(2) Avg. Discoveries before 1972

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 grouped 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. We 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.

Figure 6. **Estimated Percent Impact of Landsat for Juniors and Seniors**
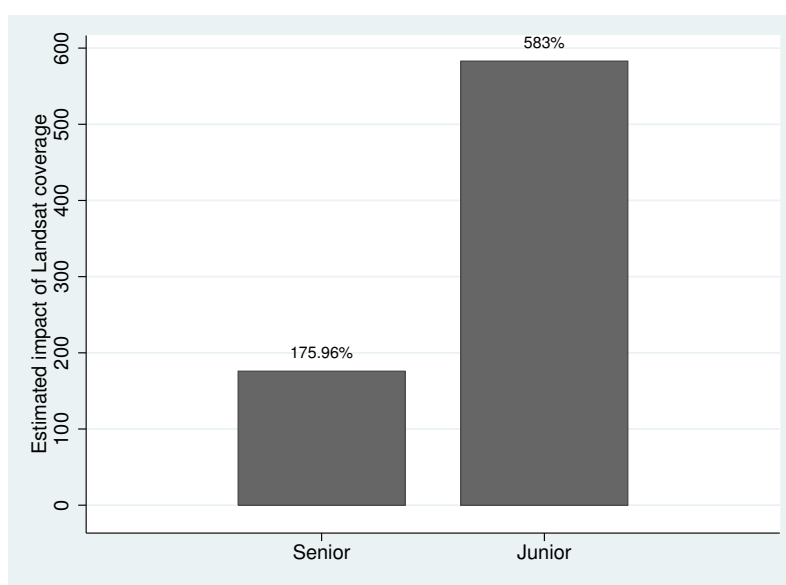
Note: This figure plots the estimated elasticities of the impact of the Landsat program separately for juniors and seniors. For each of the two estimates, the coeffcent from Table 5 is scaled by the average discoveries for juniors or seniors before the launch of the Landsat program.

Figure 7. **Estimated Percent Impact of Landsat by the Quality of Local Institutions**
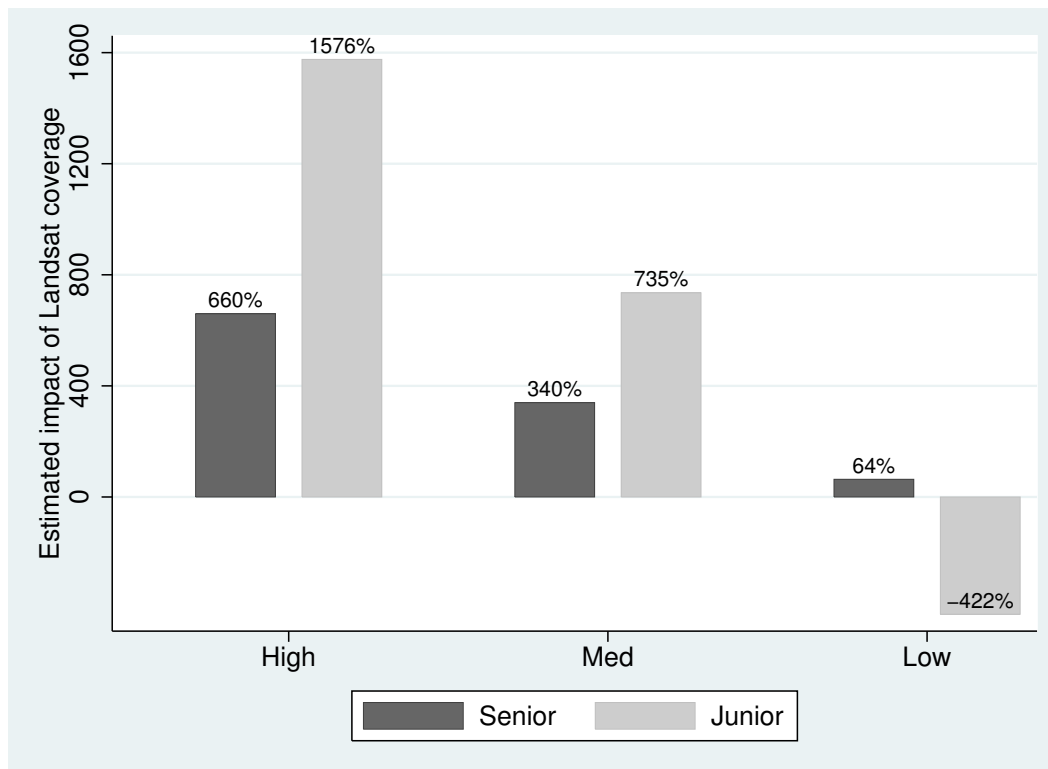
Note: This figure plots the estimated elasticities of the impact of the Landsat program separately for juniors and seniors, separately for different regions depending on the quality of local institutions. Specifically,the Fraser Institute Survey of Mining Companies ranking of 122 regions around the world in terms of "policy score" is used to split regions into terciles of "high", "medium" or "low" quality. Estimated impacts for each of these three types of regions from Table 6 are scaled by the average discoveries in corresponding regions before Landsat's launch to calculate percent changes.

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)%	4.835	21.45	0.000	0	100
1(Discovered post-1972)%	3.919	19.40	0.000	0	100
Landsat Coverage					
Landsat Intensity	45.948	57.87	20.000	0	272
Best Image Cloud-cover (%)	0.086	0.24	0.000	0	1
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	1.907	0.86	1.584	2	15

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 (excluding Antarctica and Greenland), resulting in a total of 9496 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 intensity**

Panel A – Comparison by Landsat intensity

	(1) above-median	(2) below-median	(3) diff	(4) p-val
Discoveries (pre-72)	0.023	0.015	0.008	0.04
Prospectivity Score	1.969	1.845	0.125	0.00
Publications (pre-72)	0.047	0.004	0.043	0.00
Earthquake Hazard	0.741	0.517	0.224	0.00

Panel B – Comparison by Avg. Annual Cloud Cover

	(1) Low Cloud	(2) High Cloud	(3) diff	(4) p-val
Discoveries (pre-72)	0.019	0.020	-0.001	0.81
Prospectivity Score	1.903	1.910	-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 with above-median Landsat intensity (above 20 images) with blocks with below-median Landsat intensity (below 20 images). 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 (9496 blocks for 41 years implies 389,336 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 (9496 blocks for 41 years implies 389,336 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 (9496 blocks for 41 years implies 389,336 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**

	1(Startup)		1(Incumbent)	
	Estimate	% Change	Estimate	% Change
Post X High	0.295***	1575.53	0.494***	659.81
	(0.062)		(0.104)	
Post X Medium	0.106***	735.07	0.231***	339.54
	(0.020)		(0.046)	
Post X Low	-0.022***	-422.28	0.049$^+$	62.85
	(0.008)		(0.032)	
Block FE	Yes		Yes	
Year FE	Yes		Yes	
N	309263.000		309263.000	

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

Specification:

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

Note: Block-year level observations. All estimates are from OLS models. The sample includes all block-years from 1950 to 1990 (9496 blocks for 41 years implies 389,336 block-year observations). Post: $0/1 = 1$ for a block-year after the first low-cloud image (lower than 30% cloud cover) has been received. 1(High), 1(Medium) and 1(Low): $0/1 = 1$ for blocks that belong to the first, second and third tercile of the "policy rank" distribution according to the Fraser Institute Mining Survey (see Table C.6 for complete list). 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. "% Change" is calculated by dividing the estimate by the average discoveries reported by juniors (or seniors) in High (Medium or Low) regions before the launch of the Landsat program. See text and appendix for data and variable descriptions.

8 Appendices

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.[8] 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

---

[8]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. The United States is comprised of 615 unique blocks, and all of these blocks received low-cloud Landsat imagery early in the program's history. Assuming that each of the blocks benefited from the Landsat information for about 15 years, this translates to an increased value of about $10.885 billion USD in terms of additional reserves (at a gold price of $400 per ounce). For Canada, a similar calculation yields additional resources found to the tune of $22 billion USD.

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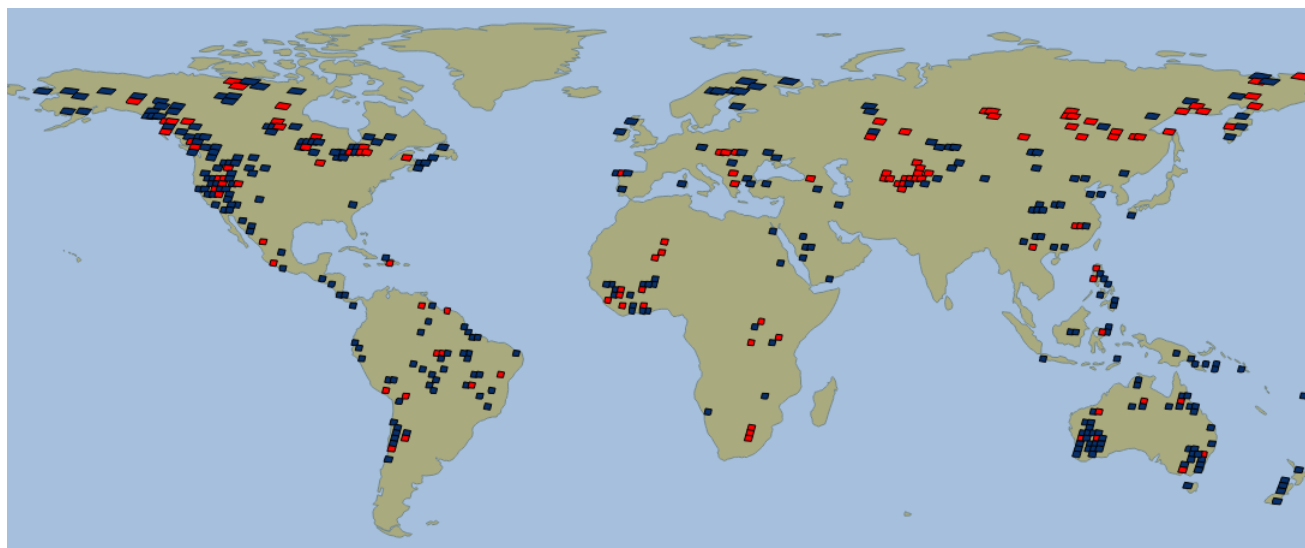
Key– 1958-1973: Red and 1973-1988: Blue

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
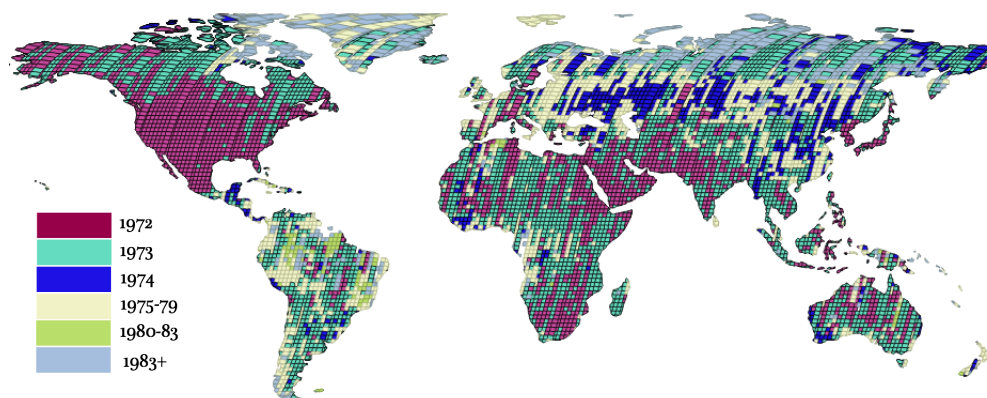
Panel B: Time-series Variation in Landsat Coverage (First Low-Cloud Image Year)
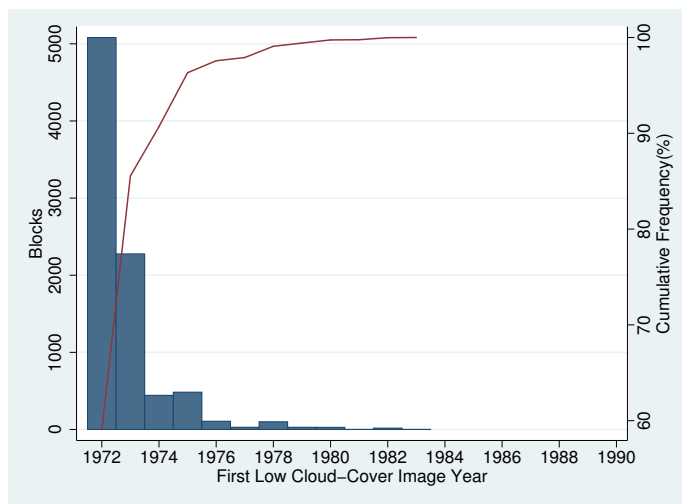
Note: This figure illustrates Landsat blocks and the variation in their mapping over time exploited in this paper. Panel A shows the location of each Landsat block, and the color represents the year in which these blocks were first mapped by the Landsat program. For blocks that were not mapped by the 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.

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

	1(Discovery)	1(Discovery)
Panel A : Excluding USA–		
N=389336		
Post Mapped	0.132***	
	(0.0305)	
Post Low-Cloud		0.154***
		(0.0277)
Panel B: No USA, Can, Aus		
N=291510		
Post Mapped	0.0953***	
	(0.0313)	
Post Low-Cloud		0.101***
		(0.0277)
Panel C : Tree-Cover Only		
N=80975		
Post Mapped	0.0172	
	(0.145)	
Post Low-Cloud		0.0803
		(0.108)
Block FE	Yes	Yes
Year FE	Yes	Yes

Note: This table presents estimates from baseline DD specification for different sub-samples of the data. Panel A excludes USA blocks, Panel B excludes blocks from USA, Canada and Australia and Panel C only includes blocks that have substantial tree cover. Tree cover is coded using codes 1-4 from the GLC2000 dataset indicating all "broadleaved" trees.

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 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.

Table C.5. **Impact of Landsat in Pre-Explored Territories**

	All Firms	Juniors	Seniors
Post Low-Cloud X Adjacent	0.554***	0.219***	0.353**
	(0.173)	(0.0642)	(0.142)
Post Low-Cloud	0.114***	0.0269***	0.0899***
	(0.0259)	(0.00650)	(0.0245)
Block FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
adj. R^2	0.00190	0.00112	0.00122
N	389213	389213	389213
Clusters	9493	9493	9493

Note: This table presents estimates of the Landsat program differentially for blocks that experienced or were adjacent to a new discovery by an incumbent firm before 1972. *Post Low-Cloud X Adjacent* codes block-year observations after a block that received a discovery before 1972 has received a low-cloud image. Column (1) includes discoveries from all firms, while column (2) includes discoveries only from juniors and column (3) includes discoveries only from seniors.

Table C.6. **Policy Rank of Regions (Fraser Institute Survey) according to Institutional Quality Terciles**

Panel A : High

Country	Region	Policy Rank
Europe	Ireland	1
Canada	Alberta	2
Europe	Finland	2
Europe	Sweden	4
Canada	New Brunswick	5
Canada	Saskatchewan	6
Canada	Newfoundland and Labrador	7
USA	Wyoming	8
Canada	Manitoba	9
Australia	Western Australia	10
Canada	Nova Scotia	11
Canada	Quebec	12
Africa	Botswana	13
USA	Utah	14
USA	Nevada	15
USA	Minnesota	16
Argentina	Salta	17
Europe	Norway	18
Australia	South Australia	19
Africa	Namibia	20
Canada	Ontario	21
Latam	Chile	22

Panel B : Medium

Country	Region	Policy Rank
USA	Alaska	23
USA	Arizona	24
USA	Colorado	25
Canada	Yukon	26
Australia	Tasmania	27
Australia	Northern Territory	28
Argentina	San Juan	29
USA	Idaho	30
Australia	New South Wales	31
Europe	Greenland	32
Australia	Queensland	33
USA	Michigan	34
Oceania	New Zealand	35
Africa	Burkina Faso	36
Canada	Nunavut	37
Canada	Northwest Territories	38
Europe	France	39
Europe	Portugal	40
USA	New Mexico	41
Canada	British Columbia	42
Argentina	Jujuy	43
Australia	Victoria	44
USA	Montana	45
Latam	Uruguay	46
Africa	Ghana	47
Africa	Morocco	48
USA	California	49
Africa	Tanzania	50
Argentina	Catamarca	51
Latam	Peru	52
Oceania	Fiji	53
Africa	Zambia	54
Asia	Thailand	55
Latam	Mexico	56
Africa	Liberia***	57
Argentina	Rio Negro	58
Latam	Guyana	59
Africa	Mali	60
Africa	Ivory Coast	61
Asia	Cambodia	62
Asia	India	63
Asia	Myanmar	64
Africa	Uganda	65
Africa	South Africa	66
Latam	Panama	67

Panel C : Low

Country	Region	Policy Rank
USA	Washington	68
Europe	Turkey	69
Africa	Lesotho	70
Africa	Mauritania	71
Latam	Nicaragua	72
Asia	Kazakhstan	73
Latam	Colombia	74
Africa	Niger	75
Argentina	Neuquen	76
Europe	Poland	77
Africa	Angola	78
Argentina	La Rioja	79
Africa	Sierra Leone	80
Africa	Eritrea	81
Europe	Greece	82
Africa	Madagascar	83
Latam	French Guiana	84
Africa	Mozambique	85
Europe	Spain	86
Latam	Brazil	87
Africa	Guinea(Conakry)	88
Argentina	Chubut	89
Latam	Suriname	90
Argentina	Santa Cruz	91
Europe	Russia	92
Asia	Kyrgyzstan	93
Africa	Democratic Republic of Congo (DRC)	94
Latam	Guatemala	95
Africa	Egypt	96
Asia	Vietnam	97
Europe	Bulgaria	98
Latam	Ecuador	99
Europe	Serbia	100
Argentina	Mendoza	101
Europe	Hungary	102
Asia	China	103
Oceania	Solomon Islands	104
Latam	Bolivia	105
Asia	Laos	106
Africa	Kenya	107
Europe	Romania	108
Latam	Dominican Republic	109
Asia	Mongolia	110
Oceania	Papua New Guinea	111
Oceania	Indonesia	112
Latam	Venezuela	113
Africa	Ethiopia	114
Africa	Central African Republic	115
Africa	Sudan***	116
Africa	Nigeria	117
Africa	Zimbabwe	118
Africa	South Sudan	119
Oceania	Malaysia	120
Oceania	Philippines	121
Latam	Honduras	122