Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry

Abhishek Nagaraj

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)
http://pubsonline.informs.org/journal/mnsc

The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry

Abhishek Nagaraj[a]

[a] Berkeley-Haas School of Business, University of California, Berkeley, Berkeley, California 94720

Contact: nagaraj@berkeley.edu, https://orcid.org/0000-0001-9049-0522 (AN)

Received: May 28, 2017
Revised: October 18, 2018; March 6, 2020; August 23, 2020
Accepted: August 26, 2020
Published Online in Articles in Advance: March 25, 2021

https://doi.org/10.1287/mnsc.2020.3878

Abstract. How does public data shape the relative performance of incumbents and entrants in the private sector? Using a simple theoretical framework, I argue that public data reduces investment uncertainty, facilitates the discovery of new market opportunities, and increases the relative market share of new entrants relative to incumbents. I shed light on these predictions by estimating the impact of public data from Landsat, a U.S. National Aeronautics and Space Administration satellite mapping program, on the discovery rates of new deposits by incumbents (seniors) and entrants (juniors) in the gold exploration industry. I exploit idiosyncratic timing variation and cloud cover in Landsat coverage across regions to identify the causal effect of public data on the patterns of gold discovery. I find that Landsat data nearly doubled the rate of significant gold discoveries after a region was mapped and increased the market share of new entrants from about 10% to 25%. Public data seem to play an important, yet relatively underexplored, role in driving performance differences across firms.

History: Accepted by Ashish Arora, entrepreneurship and innovation.
Funding: This study was funded by the Massachusetts Institute of Technology Initiative on the Digital Economy, the Kauffman Foundation Dissertation Fellowship, UC Berkeley Fisher Center for Business Analytics, and the Beyster Fellowship.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2020.3878.

Keywords: economics: microeconomic behavior • industries: mining-metals • public information

1. Introduction

In dynamic markets, both incumbents and potential new entrants need to discover new opportunities to gain competitive advantage. For example, incumbents and potential entrants must invest in discovering promising new technologies, untapped customer segments, or unexplored project opportunities to improve performance. Because such investment is expensive and uncertain, firms are increasingly relying on publicly available resources, rather than internal capital, to reduce investment costs and discover opportunities ahead of competitors (Mowery and Rosenberg 1999; Scotchmer 2004; Arora et al. 2015, 2019; Fleming et al. 2019). A large literature has studied the implications of public resources for private performance, focusing in particular on the impact of public financial capital such as grants, subsidies, and tax credits (Mansfield 1986, Lerner 2000, Agrawal et al. 2014), as well as publicly available technology from university or government laboratories that firms can use to discover new market opportunities and improve their competitive position (Henderson and Cockburn 1996, Cohen et al. 2002, Bikard and Marx 2020).

Public finance and public technology are not the only public resources that can help firms in their quest to exploit new market opportunities. Firms can also take advantage of *public data* infrastructure, which refers to the basic geographic, economic, and demographic data that governments collect and make public. Although researchers have not paid much attention to this channel, it appears firms have long made use of it. For example, in exploring the determinants of U.S. leadership in natural resources, David and Wright (1997) provide the example of how mining companies in Michigan used data from a publicly funded geological survey published in 1842 to discover lucrative copper deposits and boost firm performance. The private sector's use of big data to drive decision making has grown exponentially in the last decade, intensifying the strategic implications of public data infrastructure for the private sector (Brynjolfsson et al. 2011, Brynjolfsson and McElheran 2016).

The use of public data could not only increase performance in the private sector, but it could also influence *who* profits. Commentators have noted that public data infrastructure may level the playing field for entrants and startups by helping them capitalize on new opportunities ahead of incumbents. In fact, individual prospectors and small mining companies were among the primary group who benefited from the publication of the government survey data in Michigan in the 1840s. On the other hand, it is also possible that public data serves as another source of competitive advantage for larger firms looking to keep entrants at bay. Perhaps surprisingly, the literature on public finance and public technology has rarely studied the differential role of public support for different types of firms and has tended to focus on the aggregate implications of public resources for firms (Arora and Cohen 2015). A key question regarding the interface between public resources and private competition therefore remains open: how does public data infrastructure shape the relative performance of incumbents and new entrants in dynamic markets?

To make progress on this question, I present a simple theoretical model to evaluate the competitive implications of public data infrastructure. Building on Nelson (1982), I argue that public data provides an imperfect but useful signal about the viability of a risky opportunity to incumbents and potential entrants, thereby shaping their investment decision. Depending on the signal provided, public data could encourage firms who did not invest before to start exploring or tell firms who explored before to stop investing. Because incumbents and entrants have substantially different costs of investment, the two types of firms respond differently to the same signal. The model clarifies that public data will not always help firms discover new opportunities or help new entrants enter the market (although it could save companies money). However, given a sufficient number of firms who face high costs of investment and a sufficiently informative set of signals, public data could have a positive effect on discovery and encourage entrants, resulting in increased market share for new firms. Ultimately, whether these conditions hold, and whether the positive effects of public data on performance and entry play out in practice, remain empirical questions.

Accordingly, the heart of this study examines the empirical impact of public data on the private sector in the $5 billion gold exploration industry. Gold exploration is both expensive and risky: it requires significant capital and years of exploration to identify a gold deposit. I study the impact of public data from the U.S. National Aeronautics and Space Administration (NASA)'s Landsat satellite mapping program, which happened to contain useful geological information that clarified the value of potential gold deposits and thus had the potential to shape firms' investment decisions. I focus on the impact of Landsat satellite maps on two related dimensions: the rate of discovery of new gold deposits by exploration firms and the share of new discoveries made by entrants versus incumbents.

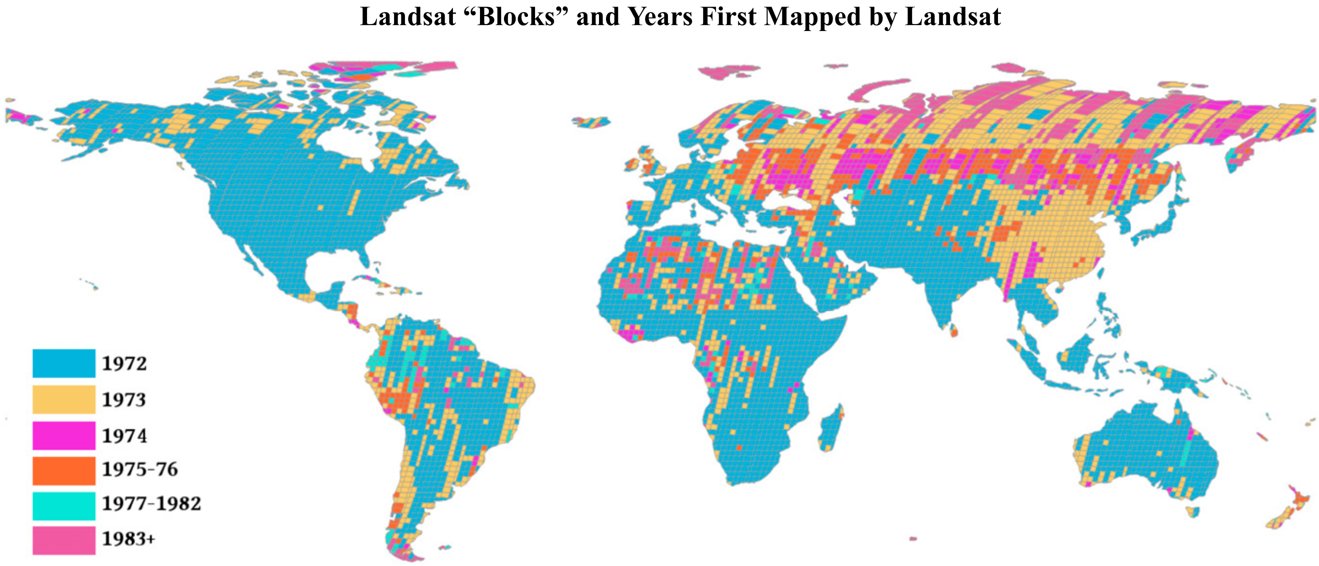
In order to examine the causal role of public data, I exploit the fact that there was significant variation in the timing of NASA's mapping effort across the approximately 10,000 blocks (regions of 100 mi^2 each) that make up the land mass on Earth. Although many blocks were mapped in the few years immediately after the program's launch in 1972, there was a long tail of regions that were mapped significantly later over the next decade. Quantitative assessments (as shown in Figure 1), and qualitative interviews indicate that though some of this variation was driven by endogenous choices (i.e., prioritizing the United States), a large part of this variation was unintentional and occurred because of technical failures and cloud cover in imagery. I combine this variation in the timing of the mapping effort with data on significant gold discoveries obtained from a proprietary database of major discoveries between 1950 and 1990. The quantitative estimates isolate the impact of the quasi-random variation in the timing of the mapping effort on exploration outcomes in a differences-in-differences framework and a battery of robustness tests help to confirm the validity of this specification.

The results of this empirical analysis suggest that public data can dramatically improve the discovery of new opportunities and encourage entry, lowering the performance advantage of incumbents. In baseline estimates, mapped regions were almost twice as likely to see a discovery than unmapped regions after controlling for region and time indicators. Furthermore, public data from Landsat maps significantly increased the share of discoveries by firms entering the gold exploration market: before Landsat, new firms made about 1 of every 10 gold discoveries; after regions were mapped by Landsat, this rate jumped to 1 in 4. Put another way, this effect translates to a 5.8-fold increase in the rate of discoveries by new entrants and a 1.7-fold increase for incumbents. Furthermore, an additional test that examines variation across regions in the effect of Landsat on gold discovery by new firms supports the mechanism that new firms might have benefited more because they face higher costs of exploration than incumbent firms. Because I do not observe costs directly, the corresponding implications for profitability are beyond the scope of this study.

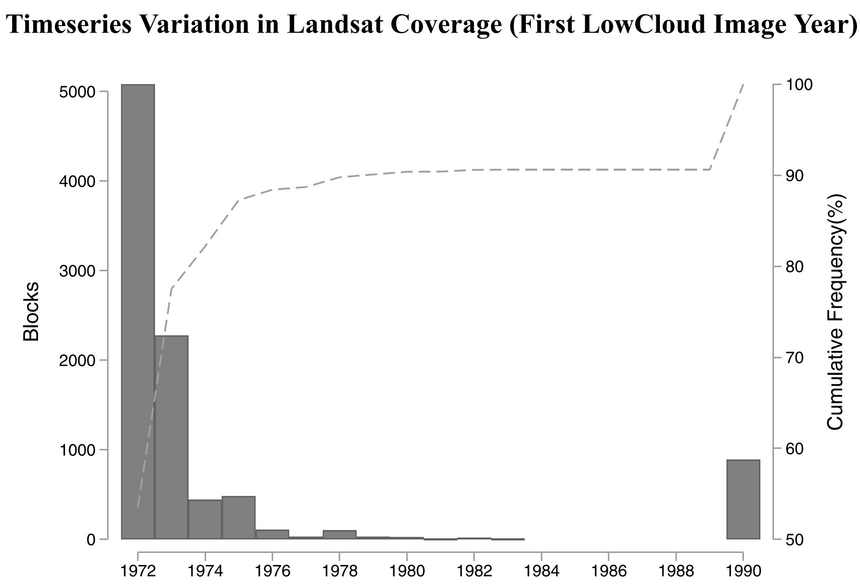
This work contributes to our understanding of the role of the public sector in shaping firm performance and market entry in markets where firms invest in risky opportunities to stay competitive. I identify an understudied channel, that of public data infrastructure, through which public resources may influence investment decisions in the private sector. I also

Figure 1. (Color online) Variation in Mapping Coverage

(a)



(b)



Notes. This figure illustrates Landsat blocks and the variation in their mapping over time exploited in this paper. (a) Location of each Landsat block, and the color represents the year in which these blocks were first mapped by the Landsat program. Blocks that were not mapped by the first phase of the Landsat program are represented in the 1983+ category. (b) Histogram for the year in which blocks were first mapped with a low cloud cover image. The frequency counts of the blocks are on the left y axis and cumulative frequency in percent is represented on the right y axis. Blocks shown to have been mapped in 1990, are in fact, blocks that I categorize as unmapped because they were not mapped by the first phase of the Landsat program, but were mapped later by following generations of satellites.

consider how public data will differentially impact two types of firms, entrants, and incumbents, based on the fact that they face different costs of investment. I explore these ideas both formally and empirically. Although satellite imagery is one type of public data, these results suggest that firms (and especially new entrants) would do well to pay attention to public data as a key strategic resource in uncertain market environments. Policy makers and governments should consider data provision as an underexploited lever to shape private competition and performance.

The paper proceeds as follows. Section 2 discusses the literature on the strategic implications of public investment and describes a theoretical framework on the role of public data in uncertain markets. Section 3 explains how the Landsat project was implemented and provides an overview of gold exploration, including and the role of incumbent and entrant firms in the industry and describes the data and research design. Section 4 highlights key empirical estimates of the role of public data in shaping discovery patterns of incumbents and entrants. Section 5 concludes.

2. Theory: Private Impact of Public Data

2.1. Public Investment and Private Performance: The Role of Public Data

How does public support shape the discovery of market opportunities in the private sector? Although the literature's focus on financial and technological channels has been insightful, it has overlooked a third form of public resource that could shape the private sector: public data infrastructure. Public data infrastructure includes any type of information about an underlying topic that is largely funded by government institutions. Three prominent examples of such information could include geographic (e.g., weather or geological maps), administrative (such e.g., individual tax records or patent applications), or scientific (e.g., the Human Genome Project, astronomical surveys, or protein databases) data (Ray 1997, Card et al. 2010, Graham et al. 2013, Marco et al. 2015, Hill and Stein 2019). Although this paper defines public data infrastructure as information being provided by governments, it could also include other sources of freely available data provided by firms (e.g., Google Trends data) or by nonprofits (such as the Wikidata project).

Much like physical infrastructure, public data forms a type of basic infrastructural investment used frequently by the private sector and especially by smaller firms (Munnell 1992, Forman et al. 2012, Agrawal et al. 2017, Seamans et al. 2017). For example, New York City's Business Atlas database has been used by retail entrepreneurs to inform location choice and raise capital (Verhulst and Young 2016). In addition, data on historical prices in agricultural markets are being used by small farmers to identify new selling opportunities (Micek 2017). Based on such examples, proponents of public data have argued that it can help firms "make better decisions," making it an "important source of economic growth, new forms of entrepreneurship and social innovation" (Ubaldi 2013, p. 4).

Despite the enthusiasm about public data, we do not yet have good theory or empirical evidence exploring how exactly public data might shape firm performance or competition. Recent work shows that publicly provided genomic information helps pharmaceutical firms discover new drugs and new applications for existing drugs (Jayaraj and Gittelman 2018, Kao 2019). Public data increase scientific productivity and helps firms patent at a greater rate (Furman et al. 2018, Nagaraj et al. 2020). However, there is much left to understand about how public information shapes the private sector. Notably, although past studies are largely focused on the aggregate implications of public data, its competitive implications remain unexplored. That there may be a differential impact of public resources on different types of firms has been hinted at in the literature but has yet to be studied rigorously. As Arora and Cohen (2015, p. 791) state, "Considerations of the effect of public support . . . have proceeded with little consideration of whether the characteristics of firms within industries may significantly moderate that effect." This paper directly investigates their theoretical idea that the impact of public resources might be heterogeneous across firm types.

In particular, what is missing from previous research is a comparison of outcomes by firm size. Many studies have shown that public resources are used at a greater rate and are more effective for smaller firms or entrants (Cohen et al., 2002, Agrawal et al. 2017, Howell 2017), but they have not compared the outcomes of smaller firms or new firms to those of larger firms or incumbents. Two recent papers compare the differential effects of public investment on private-sector patenting for larger and smaller firms. Azoulay et al. (2015) find no differential effect of public support for larger and smaller firms, whereas Furman et al. (2018) do find a differential benefit for younger companies. However, neither study looks at firms in one industry that compete with each other, and therefore it is hard to evaluate whether public information helps smaller firms gain a competitive advantage against larger ones. In sum, more research is needed to understand the implications of public support in the private sector for competition between larger and smaller firms.

2.2. A Simple Model of Gold Exploration

> "A map is not something that tells you where to go, it's a tool that lowers the risk involved in your journey. . . . [It] is the ultimate tool of investment because it derisks a process."
>
> —*Richard Jefferson, Skoll World Forum 2013*[1]

How might public data infrastructure shape the relative performance of incumbents and potential new entrants in dynamic markets? Unlike the mechanisms discussed in the literature, the benefits of public data do not seem to be technological or financial. Instead, they are likely to be informational. Specifically, public data are useful as a mechanism of triaging opportunities (Christin 2019) and guides firms' investment decisions "to proceed on a generally better set of candidate projects" (Nelson 1982, p. 462). Public data help firms distinguish valuable projects from less promising ones and increases the expected value of the investments the firm makes, thus reducing investment risk.

Furthermore, the benefits of public data are likely to disproportionately accrue to entrants. Incumbents are usually able to finance new projects internally, which reduces their cost of capital. New firms looking to enter a market are usually supported by external and more costly sources of financing. This difference means that firms facing lower costs (incumbents) are likely to invest regardless of public information, whereas firms with higher costs (entrants) are likely to invest only when the public signal is positive (Ewens et al. 2018). In the absence of a public signal, new firms might be deterred from entering the risky market at all because the expected value of the investment is too low. With a positive signal, they might now successfully enter a new market, creating competition for incumbents. Therefore, public information signals might be more valuable to smaller firms because they are more likely to change their investment behavior depending on the nature of the public signal. The following section describes a theoretical framework based on this intuition. I present a simple model (along with extensions relaxing key assumptions) that explores the conditions under which public data may increase the discovery of new opportunities and encourage entrants relative to incumbents. The empirical analysis, which follows the theoretical model, examines whether these conditions are likely to hold in practice.

2.2.1. Setup. Each firm i is evaluating a risky project (say a tract of land that may or may not have gold) that has a potential value of $V > 0$ with probability p_0 or 0 with probability $(1 - p_0)$. Firms' priors of the likelihood that their project is valuable is equal to the true likelihood, that is, p_0. Firms can invest in exploration at cost C_i and get value V or 0 depending on the state of nature. They can also not invest and get payoff 0 with certainty.

Exploration costs $C_i \in \{C_L, C_H\}$, where $C_L < C_H$. Firms come in two types S and J that have the probability that a randomly chosen firm has $C_i = C_L$ equal to either $x^S \in \{0,1\}$ or $x^J \in \{0,1\}$ with $x^J < x^S$. In other words, costs are binomially distributed with parameter x^S or x^J and a greater share of S firms are low cost than J firms. This captures the idea that entrants (e.g., juniors) usually have higher costs of investment than incumbents (e.g., seniors). The mass of firms in category S is normalized to one, so that the mass of the firms in category J is M.

Government data are essentially a set of positive or negative values for all possible blocks. I model government data to be parameterized by parameters p^+, which is the probability of discovery ($p_0 < p^+ < 1$) conditional on a positive signal and p^- ($p^- < p_0 < 1$), which is the probability of discovery conditional on a negative signal. Because beliefs are correct initially, it must be the case (by Bayes' rule) that probability of a positive signal is $q = \frac{p_0 - p^-}{p_+ - p^-}$. I allow for a map that provides positive and negative posteriors of differential strength (i.e., $p_0 - p^-$ not necessarily equal to $p^+ - p_0$). Also, positive and negative signals are imperfect: it is possible that a project has value 0 even with a positive signal (false positive) and a positive value with a negative signal (false negative).

Under this setup, firms' investment decision is decided according to a simple cutoff rule. When no public signal is available, firms invest if $C_i < p_0 V$. After the public signal, firms with a positive signal invest if $C_i < p^+ V$ and firms with a negative signal invest if $C_i < p^- V$. Therefore, a positive signal raises the investment threshold and a negative signal lowers it. Because they have different cost distributions, as I will show, the same map can have heterogeneous effects on investments across J and S firms.

2.2.2. Effect on Opportunity Discovery and Entry. I evaluate the effect of public information on the total number of opportunities discovered and the share of J firms, which is related to the total number of S and J firms who invest. Depending on the parameter values, it is possible that firms could have high (i.e., $C_H > C_L > p_0 V$) or low costs ($C_L < C_H \le p_0 V$). In the first case, no firms invest before the public signal and after the arrival of public information, firms with positive signal could invest if $C_i < p^+ V$, at least for the subset of low-cost firms. Therefore, investment and discoveries could weakly increase (decrease) when costs are high (low).

Rather than analyze these (somewhat uninteresting) cases, I will assume that costs are intermediate, that is, $C_L \le p_0 V < C_H$. The key result is that (only) under the following two conditions, the number of discoveries increases strictly and the market-share of entrants increases in lock-step as well: (a) *Entry and No-Exit:* $C_L \le p^- V$ and $C_H \le p^+ V$, or (b) *Entry outweighs Exit:* $C_L > p^- V$, $C_H \le p^+ V$, and $\frac{qp^+}{p_0} > \frac{x^S + x^J M}{1+M}$. Conversely, the number of discoveries decreases strictly if and only if $C_L > p^- V$ and $C_H > p^+ V$, or $C_L > p^- V$, $C_H \le p^+ V$ and $\frac{qp^+}{p_0} < \frac{x^S + x^J M}{1+M}$. In all other cases the number of discoveries does not change. What is the intuition behind these results?

2.2.2.1. Entry and No-Exit. In this case, before the public signal, both low-cost J and S firms invest and their high-cost counterparts do not. After public data, the negative signal does not discourage those who invested previously even if they receive a negative signal, but it encourages high cost firms with a positive signal to enter. Therefore, total discoveries strictly increase for both J and S firms. However, because J firms are more likely to be higher cost, they are disproportionately more likely to enter and explore after public information. In other words, before the signal, only the lower-cost S and J firms explore, making the market share of J firms equal to $\frac{Mx^J}{Mx^J+x^S}$. After the signal, all lower-cost firms continue to explore, but all higher-cost firms who get a positive signal explore as well. Therefore the market share of J firms increases to

$$\frac{(qp^+ + (1-q)p^- x^J)M^J}{(qp^+ + (1-q)p^- x^J)M^J + (qp^+ + (1-q)p^- x^S)}.$$

2.2.2.2. Entry Outweighs Exit. In this case, J and S firms with low costs and a negative signal exit (i.e., $C_L > p^- V$) and high-cost firms with positive signal enter (i.e., $C_H \leq p^+ V$). The overall effect on total discoveries, then, depends on two margins. For either S-led or J-led discoveries, lower the share of low-cost firms, the greater the likelihood of entry. The second factor is the structure of signals, specifically the strength of the signal p^+ and the likelihood of positive signal, relative to p_0. The intuition here is that conditional on the same number of firms investing, a stronger posterior p^+ and an increased likelihood of a posterior leads to more discoveries increasing the total number of discoveries. In fact the two factors are related, such that discoveries for each category increase if and only if x^J and x^S are under the specified threshold. In fact, total discoveries changes from $p_0(x^S + x^J M)$ to $qp^+(1 + M)$ and therefore, total discoveries increase if $qp^+(1 + M) > p_0(x^S + x^J M)$, that is, if $\frac{qp^+}{p_0} > \frac{x^S + x^J M}{1+M}$.

In terms of J market share, before the signal only the low-cost firms explore and after the signal, only firms that receive the positive signal explore. Therefore, preinformation shares are $\frac{Mx^J}{Mx^J+x^S}$, whereas the J market share after public data changes to simply $\frac{M}{M+1}$. Therefore, J-market share increases. In fact, qualitatively, I have the result that the only case where the share of higher-cost J firms decreases is if the total number is at its maximum before public information is provided (i.e., the entire mass of J firms invests) and some of these firms stop investing after a negative signal. In all other cases, the share of J discoveries increases as compared with S discoveries (see Online Appendix B for a formal derivation).

Finally, it is also interesting to consider the determinants of the share of higher-cost J firms as a function of the cost difference $\Delta x = x^S - x^J$ keeping x^S fixed. This allows us to examine the predictions of the model where J and S firms are separated only by their cost distributions (rather than on other, equally interesting dimensions such as the quality of their priors).

Specifically, in the intermediate cost case ($C_L \leq p_0 V < C_H$), if the share of discoveries by J firms goes up, then change in share of J firms is also increasing in Δx. To see why, the change in J cost share is proportional to the following expression:

$$\frac{M}{M+1} - \frac{Mx^J}{Mx^J + x^S}.$$

This term is decreasing in x^J and hence increasing in Δx. This result is driven by the fact that the increase in the J-share largely relies on higher share of low-cost firms among S-type firms. If this difference becomes larger, then the effect on the share of J firms becomes more pronounced. However, if the cost gap between junior and senior firms is so high that premapping market share for juniors is zero (i.e. the high-cost case), then we expect these results to flip. In other words, if the total amount of discoveries goes up ($C_L < p^+ V$), the share of juniors decreases in Δx ($C_H > p^+ V$) or does not depend on Δx ($C_H \leq p^+ V$). This is because either we do not get any change in the share of juniors or because change is decreasing in the difference in shares of low-cost firms.

Overall, we have the prediction that there is an inverted-U pattern of junior market share as cost gap increases between juniors and seniors. Increases in the junior cost-disadvantage initially lead to a greater benefit from mapping for this sector, but eventually this relationship becomes downward sloping. I will test this prediction in the empirical section as one test for the cost channel.

2.2.3. General Cost Distributions and Competition. The analysis thus far has assumed a binomial cost distribution and no competition between firms in the discovery of new opportunities. In Online Appendix B, I present results that relax both these assumptions. First, I evaluate the results using a general cost distribution rather than assume that costs are binomially distributed. The key insight that the overall effect of the map depends on the relative mass of firms who exit following a negative signal and those who enter following a positive signal remains valid. When relaxing the assumption about cost distributions, mathematically, the relationship between entry and exit is now affected by not only the extent of updating (i.e., $p^+ - p_0$ and $p_0 - p^-$) but also the mass of the cost distribution affected by the new signal. The key results therefore depend not just on whether the map is able to move firms' priors, but also on whether there is sufficient mass of firms in the part of the cost distribution affected by the signals from public data. In particular, I discuss the case of the uniform distribution where mass is evenly distributed for all costs between the upper and lower bound. Although in the binomial case, the results depend on x, for a uniform distribution that is *wide enough* (i.e., there are always firms who have cost high enough not to explore with positive signal, and firms who are so efficient they explore under negative signal), the effect on discoveries is always positive.

Next, I also relax the assumption that firms do not compete in the discovery process. In this extension, I assume that N firms compete to explore on K different projects. Each firm must choose whether to invest at cost C_i, and if so, on which project to invest. Multiple firms can invest in a single project. The payoff conditional on investment is $p_k V/N_k$ where p_k is the probability of discovery for a given project k and N_k is the total number of entrants for that project. Information and cost structures are the same as the baseline model. Following the literature (Bresnahan and Reiss 1990, Berry and Reiss 2007), firms engage in sequential competition, with firms who come later being able to see that history of past investments and project choices. Under this framework, I find that a sufficient condition for number of discoveries to increase is $x < (q) \times (\frac{p^+}{p_0}) \times (\frac{K}{N})$.

The intuition is similar to the baseline result: as long as the number of low-cost firms is not *too large*, public data have the potential to increase the total number of discoveries. The key additional wrinkle that emerges in this framework is that the *acceptable* share of low-cost firms also depends on K/N. The larger the number of projects relative to the number of firms, the more likely it is that public data increases discoveries. When projects are limited and the number of firms are many, the value of public data are more limited in increasing discoveries. In other words, a public map is most useful it describes a large landscape of potential projects to a relatively limited number of firms. Online Appendix B.3 provides details.

Overall, the theoretical exercise outlines that public data has the potential to increase the discovery of market opportunities and encourage entry. However, for this result to hold, public data must be informative and investment costs must be sufficiently high. In competitive markets, the number of possible opportunities in relationship to firms should be large. To examine whether these conditions hold in practice, I turn to examining the role of geological signals from NASA Landsat images on shaping the discovery of new gold deposits by seniors (incumbents) and juniors (potential entrants).

3. Empirical Setting and Data

3.1. Setting

3.1.1. Landsat Program. Landsat is the first and longest-running program to provide images of the Earth from space and offers a prime example of public data of relevance to the private sector. Starting in 1972, the Landsat program has overseen seven satellite launches that capture images of Earth with multispectral cameras. The resolution of the images does not allow for observing specific buildings or structures, but the images are useful for analyzing land use and geological features such as fractures in the surface of the earth. Each image from the first generation of the Landsat program covers an area of about 100×100 miles. In my data set, I include the 9,493 satellite images that are required to cover all of Earth's land masses (not including Antarctica and Greenland). The unit of analysis in this paper is a block of land that corresponds to a Landsat image.

The focus of this paper is the first generation of satellites in the Landsat series (Landsats 1, 2, and 3) that operated between 1972 and 1983. It was not possible for NASA officials to significantly change the orbits of these satellites; however, program operators usually controlled what locations were prioritized for data collection through regular instructions issued to the satellites. The Landsat satellites orbited the surface of the earth every 18 days, so in principle, it was possible to capture an image of every location on earth every 18 days. In practice, however, the cameras did not operate at all times because of technical and operational issues (Goward et al. 2006, p. 1155), and many regions were left unmapped for almost a decade after the launch of the program because of difficulties with collecting, storing, and relaying data back to NASA. I discuss the timing of Landsat images in more detail in Section 3.3.

The photos taken by the Landsat satellites were relayed to the Earth Resources Observation and Science (EROS) center in Sioux Falls, South Dakota. The center then distributed the data as tapes or physical images at a reasonable cost and without intellectual property considerations, as required by law. The prices for the data ranged from about \$10 for a 10-inch negative to about \$50 for a 40-inch color photograph (Draeger et al., 1997). Because all Landsat imagery was collected at EROS, by studying the archives of this institution, I am able to collect information as to the location of each block, when each block was imaged, and the quality of the images, including a measure of cloud-cover at the image level.[2] According to one estimate, the cost of the program when it was launched in 1972 was approximately \$125 million (Mack 1990, p. 83).

3.1.2. The Gold Exploration Industry. Gold is the next most intensively explored natural resource after oil and gas, and gold mining is a capital- and time-intensive process. Even though satellite imagery is helpful for all types of mining, the gold mining industry is one of the largest in this field and allows for a clear comparison between seniors and juniors. Seniors are large incumbent firms that both operate mines and invest in exploration and juniors are small entrant firms mostly funded by risk capital that are purely in the exploration business (Humphreys 2016). Junior firms are treated as market entrants because they are looking to make their first (and usually only) major discovery and thereby enter the market. A successful junior will return the proceeds to their investors and founders through sale or acquisition of the assets and so it is rare for a junior to transition to a senior, although some juniors have made this transition. Government geological agencies are also involved in gold exploration, and given their relatively large size, they will be treated as part of the seniors group.

Seniors and juniors explore for gold through a risky two-step process: initial triaging using remotely sensed information sources such as satellite imagery followed by more expensive on-the-ground exploration and drilling. For a given target, a key decision for a firm is to use cheaply available information in the first step to decide whether to invest more costly resources in follow-on exploration. The key question is whether access to public data in the first step changes the relative discovery rates for juniors and seniors.

This two-step process is costly, and juniors face disadvantage in this regard relative to seniors. While seniors usually fund exploration through internal sources, juniors rely on more costly sources of external capital (e.g., private placement or equity financing) (Humphreys 2016). Juniors also face higher costs of access to physical assets (such as drilling and survey equipment) because they are not able to spread the costs of this equipment across projects like senior firms. Juniors also face higher transaction costs (e.g., obtaining property rights, environmental assessments, and other permissions from local governments) relative to seniors because they fewer have connections and lower experience with these activities (Schodde 2014).

Not only do juniors face higher costs of exploration as compared with seniors, these differences also vary by geography. In particular, this difference increases in regions with poor institutions because uncertain property rights and the resulting risk of expropriation increases exploration costs. For example, the American junior firm Mundoro Mining that was operating in China for over five years had to leave the country after discovering promising prospects because government agencies would not grant them the relevant business and exploration licenses (Hart 2014, p. 135). Such incidents are common with junior firms, and therefore investors increase their cost of capital seeking additional return for investing in regions where property rights are less assured. Higher capital costs and increased expropriation risk means that the cost difference between juniors and seniors increases in regions with poor institutional quality.

The arrival of Landsat imagery in the early 1970s had important implications for the gold mining industry and techniques to use satellite imagery to guide gold exploration were discovered (Rowan 1975, Vincent 1975, Rowan et al. 1977, Krohn et al. 1978, Ashley et al. 1979). Landsat imagery allowed geologists to spot geological features, such as faults and lineaments, that might otherwise have gone unnoticed. Accurate knowledge of faults and lineaments is crucial for geologists because mineral resources often occur along these features. Landsat was the only satellite imagery provider in this period, and although far from perfect, it was an important tool for firms to reduce uncertainty in their exploration process.[3] Although it was possible to use airplanes to collect aerial imagery (Spurr 1954), this process was expensive.

In the context of the theoretical framework, Landsat information can be thought of as providing positive signals when certain geological features associated with gold deposits are present and negative signals when such features are absent. Although firms had a general idea of the overall probability of discovery for a set of projects (p_0), Landsat can be modeled as providing a negative signal for a majority of projects, but increasing expected probability of discovery for a small number of promising ones. If many junior firms were not investing because of high exploration costs and were induced to invest as a consequence of a positive signal, we might expect Landsat data to encourage discovery and entry. Although the use of Landsat information is widespread and textbooks claim that "providing basic geoscientific information by governments can act as a catalyst for mineral exploration in unexplored areas," the empirical effects of Landsat information on gold exploration remain unexamined (Gandhi and Sarkar 2016, p. 177).

3.2. Research Design

Although Landsat information was useful for the gold exploration industry, it was not available for all regions of the world at the same time. Instead, there was significant variation in time and space across which Landsat data were collected and made available. Before using this variation to examine the effects of public data on gold exploration, it is important to understand the drivers of this variation. Although cross-sectional variation linked to a region's potential for gold is controlled for via block fixed effects, the concern still remains that gold exploration firms are shaping the trajectory of satellite mapping. If firms lobby NASA to increase coverage in regions of interest to them, reverse causality would be a significant issue in the research design.

However, investigations of variation in coverage by Landsat experts suggest that rather than lobbying, timing variation is 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). In fact, this variation was both unexpected and unnoticed until quite recently. An interview with a Landsat administrator confirmed, "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" (Interview, April 8, 2015). The Landsat administrators I interviewed also said that the Landsat planning team was deliberately insulated from firms in the private sector (such as exploration companies) because NASA did not want to be seen to be catering to the needs of a select few. Beyond technical limitations, variation in coverage was also due to cloud cover, which remains a central challenge in the remote sensing industry. Cloud-cover is unpredictable and hard to work around as indicated by one of my interviewees: "Our ability to predict clouds [is limited] ... after a few tries you might end up with only about one or two scenes that are very clear" (Interview, November 22, 2014).

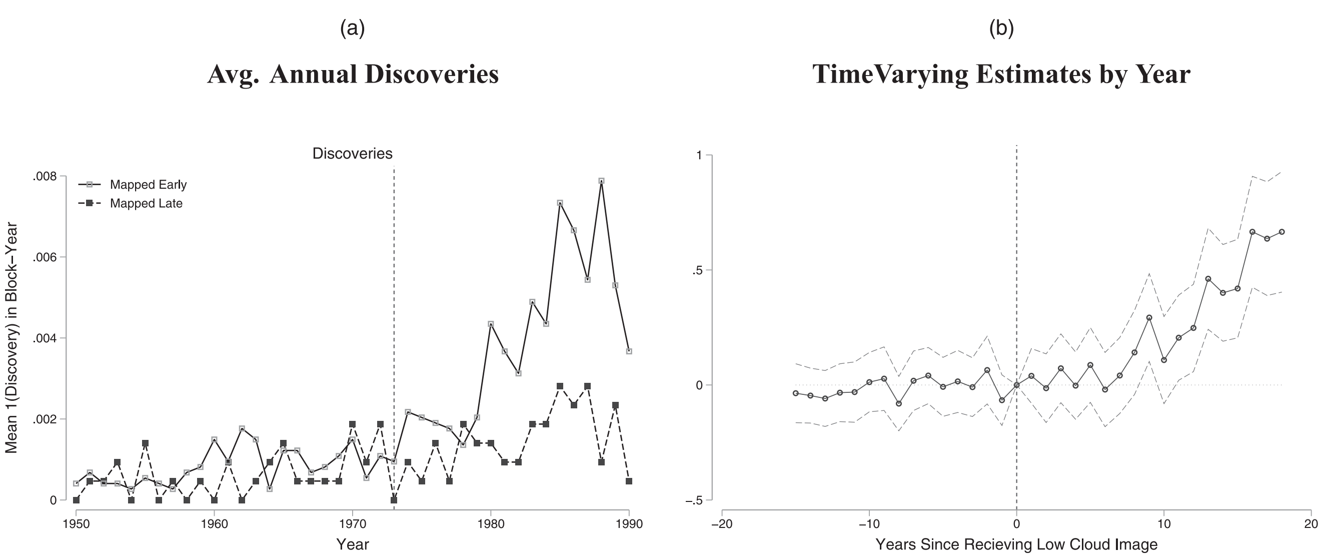
Beyond this qualitative validation of the research design, Figure 1 provides a map showing the timing of the mapping effort across blocks around the world (a), as well as a histogram of the years in which blocks were first mapped by the Landsat program (b). This evidence makes clear that (i) there is significant variation in terms of the geographic location of blocks that were mapped early or late (although there is significant clustering, especially in the United States), and (ii) whereas a majority of the blocks were mapped for the first time in the first two years of program operation, there is a long tail of blocks that were mapped considerably later in the program. A simple time-series comparison of average gold discoveries between blocks that received Landsat coverage early (mapped 1972–1974) versus late (after 1974) confirms this finding (Figure 2(a)). Rather than showing that blocks mapped early were in regions known for their potential for gold, this figure suggests no correlation between average discoveries and the timing of mapping: discoveries in blocks mapped early and late had fairly flat and parallel growth rates before 1973, when Landsat data were first made available.

3.3. Data

Having built some confidence in the validity of the research design, I now turn to describing the data needed to explore the relationship between Landsat maps and the discovery of new gold deposits. The data set I build is at the block level and (a) quantifies the timing and spatial variation in Landsat coverage, (b) records gold discoveries by juniors and seniors, and (c) (to examine the cost channel) provides indicators for cost differences between juniors and seniors.

3.3.1. Landsat Coverage Data. I construct a data set of the timing and spatial variation of Landsat images from the EROS data center's sensor metadata files.[4] These data help me construct my main independent

Figure 2. Impact of Landsat on Gold Discoveries over Time



Notes. The figures depict the impact of Landsat imagery on gold discovery over time. (a) Average discoveries over time of blocks mapped early (1972–1974) and blocks mapped later (after 1974). The vertical line at 1972 shows when the Landsat program commenced operation. (b) Estimates and 95% confidence intervals of β_t from the event study specification in Section 4.1. This panel describes the estimated difference between treatment and control blocks for years relative to a year zero, which marks the year when a block first received low-cloud imagery.

variables by providing a list of all images collected by the Landsat cameras, including the location being imaged, the date the image was collected, and information about the quality of the image, including an assessment of cloud coverage in the image (Goward et al. 2006). 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. The data were made available for follow-on use immediately after they were available at the EROS data center, and there were no significant delays in disseminating this information to downstream users. Second, I construct the variable $Post\ Low - Cloud_{it}$, which is an indicator variable expressing whether a block's image has sufficiently low cloud coverage to be useful, defined by experts as less than 30% cloud cover.

3.3.2. Measuring Discoveries. It is a nontrivial exercise to collect data on all gold discoveries by junior and senior firms because there is no official 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 between 1950 and 1990 (Schodde 2011). These data have been collected over a period of 15 years using press reports, disclosure documents, and other industry sources. I match the location of each discovery to a specific block. Since multiple discoveries in the same block-year are extremely rare, my key dependent variable is *Any Discovery*$_{it}$, which is an indicator variable for whether a discovery was made in a given block-year. In total, 460 unique blocks have seen a total of about 740 significant discoveries between 1950 and 1990 (as shown in online appendix Figure D.1). Furthermore, 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. The data provider lists a firm as a junior if it is primarily involved in exploration and is looking to enter the industry since it does not have any operating mines, while seniors are incumbents who already operate existing mines. Online Appendix A provides more details on data construction, how discoveries are defined, and how the date of discovery is coded.

What counts as a *new discovery* is a matter of significant debate among geologists because some sites could be rediscovered, whereas nondiscoveries could later turn out to be meaningful (Slade 2001). Furthermore, discoveries might be under-reported in certain regions (e.g., erstwhile USSR or China) or might be exaggerated or even falsified (Brown and Burdekin 2000). Although it is difficult to completely address this challenge, selective discovery does not pose a serious threat in this setting because (a) a cross-validation exercise with comparable data sets shows that 93% to 99% of all valuable discoveries are included (Schodde 2011), (b) the results are robust to using only large discoveries (more than 2.2 million ounces of proven reserves) where these challenges should be more limited, and (c) because I define discovery as the date of the first economic drill intersection which is a conservative method to define a discovery. Finally, as I will demonstrate, my results are robust to switching to an alternate source of discovery data that I purchased from SNL Metals and Mining, although these data are available only for the later years in my study period.

3.3.3. Measuring Cost Variation. Next, I collect data to explore whether the change in junior market share in the presence of public data stems from their cost disadvantage. In particular, I rely on the Survey of Mining Companies, conducted by the Fraser Institute (Jackson 2014) in 2014, which measures institutional conditions in the mining industry to obtain a relatively good measure of the institutional barriers to operating in a region or country. As described in Section 3.3.2., my assumption here is that these barriers increase project costs more for junior firms than for senior firms because senior firms are able to overcome institutional and regulation-related barriers at a lower cost. Responses are collected from surveys administered to more than 4,200 managers in the industry who are asked about costs of exploration arising from institutional features such as environmental regulation, legal institutions, labor regulations, and so on. I use this survey to categorize blocks in jurisdictions that rank below the median on this dimension in this survey as places where juniors are likely to have larger costs as compared with seniors.

3.3.4. Summary Statistics. Table 1 lists the key variables used in the quantitative analysis with the 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 1 if a new gold discovery is reported in a block-year. This variable is scaled by a factor of 100 for legibility throughout the analysis. The mean of this variable is 0.188, which means that there is a 0.188% chance that a discovery is reported in a block-year. *Any Junior Disc* is set to 1 when *Any Discovery* takes the value of 1 and at least one discovery was reported in a block-year by a junior firm. On average, 0.038% of block-year observations had 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 1 if a block has been mapped or mapped with a low-cloud image, respectively,

Table 1. Summary Statistics

Variables	Mean	SD	Median	Minimum	Maximum
Panel A: Block-year level					
Outcome					
Any discovery (%)	0.188	4.33	0.000	0	100
Any junior discovery (%)	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
Panel B: Block level					
Outcome					
Total discoveries	0.083	0.52	0.000	0	16
Total junior-led discoveries	0.017	0.18	0.000	0	7
1(ever discovered)%	4.846	21.47	0.000	0	100
1(discovered post-1972)%	3.940	19.45	0.000	0	100
Landsat coverage					
Year first mapped	1,973.222	3.58	1,972.000	1,972	1,990
Year first low-cloud	1,974.368	5.19	1,972.000	1,972	1,990

Notes. Blocks correspond to Landsat images that divide the planet in blocks of approximately 100 × 100 miles. All blocks that cover the earth's landmass are included and water bodies (as well as Antarctica and Greenland) are excluded, resulting in a total of 9,493 blocks in my sample.

by the Landsat program. There is a small percentage of blocks that were never mapped by the first generation of the Landsat program. For these blocks, the *Post Mapped* and *Post Low − Cloud* variables are always set to zero, although the results are robust to excluding these blocks altogether (online appendix Table D.5B). Table 1, panel B, provides summary statistics for variables that vary across blocks but not over time. These data indicate that about 4.8% of the blocks reported a discovery between 1950 and 1990 and about 3.9% of blocks reported a discovery after 1972, the year when Landsat was launched. These data also show that the median block was mapped by a low-cloud image in 1972; however, there is a long tail of blocks that remain unmapped until 1990.

4. Results

I now turn to analyzing whether Landsat maps boosted the discovery of gold and whether any potential increases accrued differentially to juniors as compared with seniors.

4.1. Did Landsat Increase Gold Discovery?

I use Ordinary Least Squares (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. Y_{it} indicates either the total number of discoveries or the number of discoveries by senior or junior firms in a given block-year. The term $Post_{it}$ represents either $Post\ Mapped_{it}$ or $Post\ Low - Cloud_{it}$, which equal to one for a block after it has been mapped or it has been mapped with a low-cloud image, respectively. 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 are mapped earlier by Landsat do indeed report more gold discoveries and earlier discoveries than blocks mapped later, then I should find that the difference-in-difference estimate β_1 is positive. All my specifications cluster standard errors at the block level to address the concern that discoveries within blocks are likely to be correlated over time. In additional robustness checks, I include more general clustering (e.g., at the country and block-group levels) that takes seriously spatial proximity between different blocks. I find that the results are generally robust to these additional restrictions.

Table 2 presents estimates from this regression for Total Discoveries for both the $Post\ Mapped_{it}$ and $Post\ Low - Cloud_{it}$ variables without (columns 1 and 2) and with block fixed effects (columns 3–5). The coefficients generally reduce in size after controlling for block fixed effects, indicating their importance in this setting. The main result is that, after controlling for block- and year-level fixed effects, there is a positive impact of Landsat images on gold discovery suggesting that public data can increase private performance. Specifically, the estimate of β_1 indicates an average increase of between 0.152 and 0.164 percentage points on the likelihood of a gold discovery after a Landsat image becomes available, a significant increase given that the baseline rate of discovery is about 0.19%. This means the rate of discovery in imaged regions is almost doubled, albeit on a low base-rate. Column 5, which includes both $PostMapped_{it}$ and $Post\ Low - Cloud_{it}$, is particularly

Table 2. Baseline Estimates for the Impact of Landsat on Gold Discovery

	Any discovery				
Variables	0.251***		0.152***		0.0178
Post mapped	(0.0265)		(0.0294)		(0.0399)
Post low-cloud		0.267***		0.164***	0.155***
		(0.0276)		(0.0274)	(0.0370)
Block fixed effects	No	No	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
N	389,213	389,213	389,213	389,213	389,213

Notes. Standard errors clustered at block-level shown in parentheses. Block-year level observations, estimates from OLS models. The sample includes all block-years from 1950 to 1990 (9,493 blocks for 41 years equals 389,213 block-year observations). *Post Mapped*: 0/1 = 1 for a block-year after the first image has been received and *Post Low – Cloud*: 0/1 = 1 for block-year after the first low-cloud image (less than 30% cloud cover) has been received. 1(Discovery): 0/1 = 1 if a discovery is reported in a block-year.
$+p < 0.15$; $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

interesting because the coefficient on $Post\ Low - Cloud_{it}$ variable is 0.155 and significant, whereas the estimate on the $Post\ Mapped_{it}$ variable is small and not statistically different than zero. This suggests that the effect of the Landsat mapping depends on its information content (rather than any other channels).

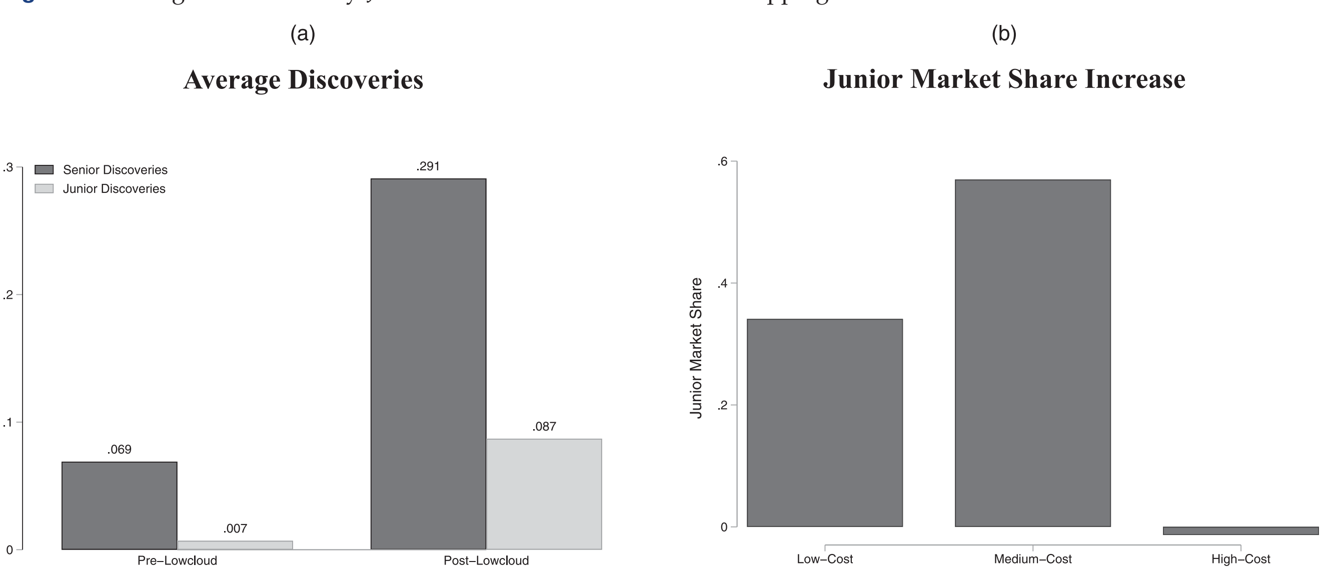
4.2. Junior vs. Senior Discoveries

Taking the theoretical model seriously, the baseline results indicate the conditions needed for Landsat images to lead firms to increase discovery are likely to hold in practice. Landsat imagery seems to have been informative in moving firms' assessments of potential targets. And perhaps more importantly, gold exploration seems to be sufficiently costly that the number of nonexploring firms incentivized to explore following a positive signal seems to have outweighed the number of exploring firms who might have stopped exploring after a negative signal. I now test the prediction that junior firms are more likely to increase their market share as compared with lower cost incumbents.

As a preliminary exercise to examine whether Landsat benefits junior firms more than seniors, Figure 3(a) compares the average number of discoveries by juniors and seniors before and after the arrival of a low-cloud image. Before public mapping, juniors are a very small part of the industry, making an average of only 0.007 discoveries per block-year, whereas seniors make about 0.069 discoveries per block-year. However, these proportions change drastically post-mapping. Both seniors and juniors increase their rate of discoveries significantly, but juniors have a much greater increase. Specifically, juniors now make an average of 0.087 discoveries per block-year while seniors make about 0.29. In other words, junior market share increases from about 9.2% to about 23%.

I estimate regressions similar to the baseline specification to formally test this pattern. I first set the dependent variable equal to 1 if junior firms made a discovery and 0 otherwise (Table 3, columns 1 and 2) and then set the dependent variable equal to 1 if senior firms made a discovery and 0 otherwise (Table 3, columns 3 and 4). The estimates of β_1 from these regressions provide separate estimates of the increase in

Figure 3. Average Discoveries by Juniors and Seniors Pre- and Post-Mapping



Notes. (a) Average discoveries for juniors and seniors for all block-years before and after a block is mapped by a low-cloud image. (b) Estimated junior market shares using the baseline specification across three kinds of blocks based on cost differences between juniors and seniors as proxied by a measure of institutional quality. Regions that rank highly on this measure (first quartile) comprise the low-cost group, regions in the middle (second quartile) are the medium-cost group, and below-median regions are the high-cost group.

Table 3. Impact of Landsat on Gold Discovery for Juniors vs. Seniors

	1(Junior)	1(Junior)	1(Senior)	1(Senior)
Variables	0.0288***		0.127***	
Post mapped	(0.00563)		(0.0285)	
Post low-cloud		0.0472***0000		0.121***
		(0.00651)		(0.0260)
Block fixed effects	**355.68%**	**583%**	**182.39%**	**174.95%**
Year fixed effects	Yes	Yes	Yes	Yes
N	Yes	Yes	Yes	Yes
Post mapped	389213	389213	389213	389213

Notes. Standard errors clustered at block-level shown in parentheses. Block-year level observations, OLS models. Post mapped: $0/1 = 1$ after the first image and Post low-cloud: $0/1 = 1$ after first low-cloud image. 1(Junior): $0/1 = 1$ if for junior discovery and 1(Senior): $0/1 = 1$ for senior-led discovery.

$+p < 0.15$; $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

discovery by juniors and seniors, facilitating a comparison of whether Landsat helped one group more than the other.

Estimates in Table 3 suggest that the impact of the Landsat program on juniors is about 0.047 while the impact for seniors is about 0.12. In other words, the total gain from the Landsat program (about 0.16 percentage points more) is divided such that smaller firms make 0.04 percentage points more discoveries per block-year while seniors capture the remaining 0.12 percentage points. In terms of percentage points, it seems then that seniors benefit more than juniors from the Landsat mapping effort. However, when the previous market share of juniors is taken into consideration, this interpretation changes considerably. Specifically, before the Landsat program was launched, the probability that a junior firm would make a gold discovery in a given block-year was just 0.008%, whereas for seniors it was 0.0694%. This suggests that seniors were almost entirely responsible for new gold discoveries prior to the Landsat program. After the arrival of Landsat images, however, juniors made one out of every four discoveries. Thus, it seems that the Landsat program encouraged the entry of junior firms in this industry by allowing them to make new discoveries. Put another way, juniors were 5.8 times more likely to report a discovery in mapped regions than in unmapped regions, while seniors only benefited by a factor of 1.7. 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 enter and make considerable gains in performance.

4.3. Robustness Checks

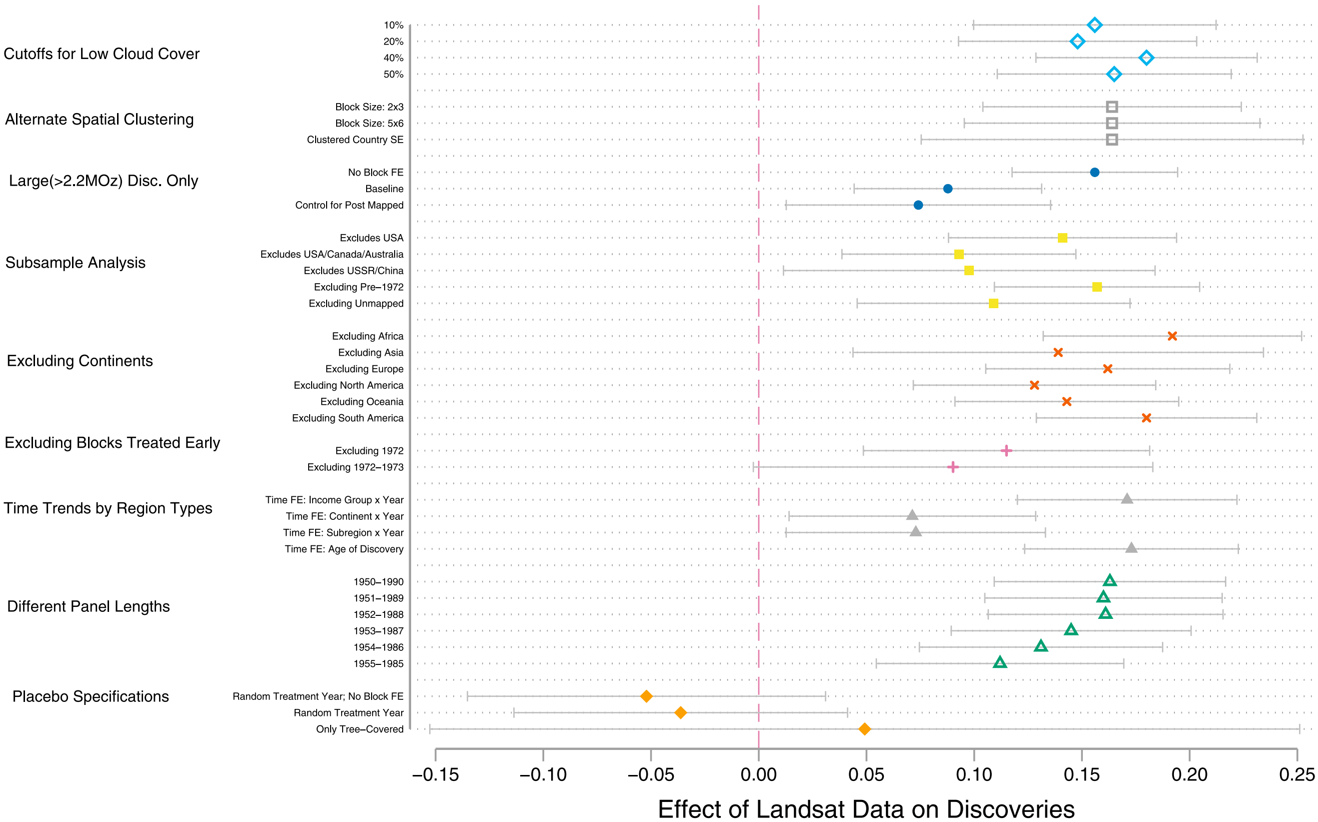
The central result of this study around increased total discoveries and increase junior market share rely on a number of assumptions about the validity of the baseline specification. I investigate a number of these assumptions in detail. I test the robustness of the design with the baseline estimates on total discoveries, and then provide additional robustness checks for the results specific to junior and senior firms.

4.3.1. Time-Varying Estimates. First, I estimate the time-varying impact of Landsat coverage on gold discovery in order to examine whether preexisting differences in gold discovery trends between recently mapped and soon-to-be-mapped places do not drive the key outcomes. 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 number of years that have elapsed since a block was first mapped with a low-cloud image. For the small percentage of blocks that never get a low-cloud image, z is always set to zero. Figure 2(b) presents estimates of β_t from this regression, which measure the difference in the number of discoveries between imaged and non-imaged blocks for every lag year. This figure makes two points. First, there are no preexisting differences in gold discovery trends between blocks mapped early and those mapped late. Second, there is a large and persistent increase in discoveries that appears after about seven to eight years after a low-cloud image. This delay accords well with my interviews with personnel at gold exploration companies, who confirm that Landsat images contribute to early-stage exploration and are typically followed by years of further exploration before a gold discovery occurs.

4.3.2. Instrumental Variable and Cross-Sectional Specifications. First, in addition to the baseline specification, I also investigate a set of specifications that use cloud-cover as an instrument for the timing of Landsat mapping. I describe this strategy and present the results in Online Appendix C. Overall, the instrumental variable analysis confirms the baseline specification, although these results are considerably larger. Although the IV provides a useful check on the baseline estimates, I emphasize the baseline estimates, which are more conservative and where the identifying variation is more transparent.

Second, I also implement a cross-sectional specification that does not rely on the panel variation within blocks. This specification under-emphasizes small timing differences in the arrival of maps and instead

Figure 4. (Color online) Evaluating the Robustness of the Research Design



Note. This figure explores how the baseline estimate on the *Post Low – Cloud* variable changes in the baseline specification for a variety of alternate specifications. Online Appendix D explains each of these alternate specifications and presents the underlying regressions that provide the estimates for this figure.

simply relates overall delays to the probability that any discovery is reported in the 20-year period after the launch of the Landsat program. The baseline specification takes the form $Y_i = \alpha + \beta_1 \times Delay_i + \gamma_i + \epsilon_i$, where the main outcome variable, Y_i, is an indicator for whether any discovery was made in a given block i between 1972 and 1990, $Delay_i$ is the difference between the year in which a block was mapped with a low cloud image and 1972, and γ_i represents spatial fixed effects, such as at the continent, subregion, or block-group level. The estimates are presented in online appendix, Table D.12. This specification is also helpful because I can evaluate the robustness of the baseline results to using an alternate source of data for the dependent variable, SNL Metals and Mining. The SNL data have limited coverage before 1982 and therefore cannot be used in the main specification but can be used here as an alternate measure of gold discovery. I also estimate another version of the cross-sectional specification where the dependent variable is time from 1972 until the first discovery in a given block. In all the three cross-sectional specifications, the estimates are statistically significant, indicating that greater delays in Landsat mapping are associated with a greater delay in the discovery of deposits.

4.3.3. Additional Robustness Checks. In addition to examining the concern around preexisting trends and using the IV and cross-sectional specifications, a number of additional concerns are worth examining as well. I estimate alternate models to test for these concerns and provide a consolidated figure with the coefficient for the $Post\ Low - Cloud_{it}$ variable in Figure 4. I provide a brief description of these concerns later.

First, the *Post Low – Cloud* variable uses 30% as the cutoff for a low cloud image. I investigate robustness to this cutoff. Second, given spatial auto-correlation between blocks, I examine robustness to clustering standard errors among geographically proximate blocks. Third, given concerns about mismeasurement in the dependent variable, I restrict my focus to only significantly large discoveries (above 2.2 million oz. of proven reserves), where missing data are less likely to be of concern.

Fourth, the spatial variation in the blocks that were mapped early is not completely random. In particular, the United States seems to have been mapped early, whereas parts of the USSR seem to have been mapped late, although variation in other parts of the world seems more idiosyncratic. In order to address concerns that regions with problematic variation are driving the effect, I repeat the analysis for different subsamples of the data excluding just the United States; excluding the United States, Canada, and Australia (the top three gold producers in the world); and excluding the USSR and China (where measurement error of gold discoveries is likely to exist). It is also helpful to exclude subsamples of blocks that might be viewed as problematic. I therefore repeat the analysis excluding blocks where discoveries had already been reported prior to 1972 and excluding blocks that were never mapped by Landsat.

Fifth, I examine geographic variation in the baseline results in order to verify that no one region is driving the overall results. I estimate the baseline specification six times, each time excluding blocks from one of the six continents I studied. Sixth, as Figure 1 indicates, the timing of the mapping is such that approximately 75% of the globe was mapped in the first two years of the Landsat program, after which there was a significant delay before the remaining blocks were mapped. While this pattern is not a direct concern for the analysis in terms of identification, I estimate the baseline prediction excluding blocks mapped in 1972 and both 1972 and 1973.

Seventh, there is a concern that Landsat mapping coincides with improvements in institutional quality in a number of major regions (e.g., the former Soviet Union), and so confounds the direct effect on gold exploration. To account for this concern, I present estimates that provide region-specific trends using three different ways to group blocks into regions. I also present estimates by including a fixed effect for the number of years since a discovery has been made in a given block. Eighth, given that the choice of start and end dates for the panel (1950–1990) is somewhat arbitrary, I present estimates by differing length of the panel.

Finally, I present estimates using *placebo* treatments to make sure that the results are not a mechanical artifact of the research design or measurement process. To do so, I randomly assign it a year between 1972 and 1990 in which the block was *mapped* and run the analysis using this *fake* treatment year. Furthermore, my interviews with executives in the gold industry revealed that, although Landsat is useful for understanding the geology of a region for gold exploration, its utility is severely diminished in regions where tree cover obscures details of the land surface underneath. Accordingly, I use a data set of global tree cover to extract blocks that contain significant tree cover and estimate the impact of Landsat on this limited group only.

Figure 4 presents estimates for each of these eight sets of robustness checks. As is clear from this chart, the baseline finding that Landsat images increases total discoveries is corroborated across these specifications, although the magnitude of the estimate does change somewhat depending on the specification. Online Appendix D provides the estimates and tables underlying this chart.

4.3.4. Robustness: Juniors vs. Seniors. Now, it turns to evaluating a few concerns that relate more directly for the result that juniors increase their market share following the availability of Landsat information.

First, online appendix, Figure D.6, estimates the event study specification similar to Figure 2. As these charts show, both junior and senior firms have a flat pretrend and discoveries grow only after the mapping. Second, I estimate robustness checks that evaluate whether the differential benefit for junior firms holds when excluding certain key regions and when counting large discoveries. Online appendix Table D.15 provides these estimates.

A specific concern that arises with the result that speaks to the impact on juniors is that seniors might be *outsourcing* their exploration to juniors, rather than losing out to them per se. Therefore, in online appendix Table D.13, I use data on joint ventures to test this idea. I find that while outsourcing could be relevant, most of the junior firm discoveries represent discoveries from capital-constrained firms. Overall, these robustness results are in line with the theory that juniors benefit more from the arrival of public data.

4.4. Probing the Cost Mechanism

In Section 2.2, I derived the result that as the project costs for juniors increase, the increase in their market share initially increases (intermediate cost case), but beyond a certain threshold it decreases (high cost case). In other words, change in share of juniors increases with increasing Δx and is then decreasing. I now test this idea empirically. Finding results consistent with this pattern would be evidence of the validity of the causal mechanism in the theoretical model, which identifies the difference in the project costs as one important reason why public information infrastructure is particularly beneficial for junior firms.

In order to test this prediction, I rely on data from the Fraser Institute Survey of Mining Companies on the quality of local institutions (Jackson 2014) as a proxy for differences in cost between juniors and senior firms, as described in Section 3.3.3. The Fraser survey provides a rank of institutional quality for 122 national and subnational regions around the world. Regions in the first quartile of the Fraser rankings have the highest quality institutions and are thus considered low cost; regions in the second quartile are medium cost; and regions in the third and fourth quartiles are considered to be high cost. I divide the countries in this way because most of the meaningful variation in institutional quality is found among the top 70 or so regions (the first and second quartile), given that there is a long list of regions with poor institutions and low levels of gold exploration. All regions below the median are therefore classified as high-cost regions. Using this measure, I modify the baseline specification to include separate coefficients for the effect of mapping for each of these three cost categories. Specifically, I estimate $Y_{it} = \alpha + \beta_1 \times Post_{it} \times LowCost_i + \beta_2 \times Post_{it} \times MediumCost_i + \beta_3 \times Post_{it} \times HighCost_i + \gamma_i + \delta_t + \epsilon_{it}$, where β_1, β_2 and β_3 are the coefficients of interest.

Table 4 presents results from this specification using both the $Post\ Mapped_{it}$ and $Post\ Low - Cloud_{it}$ variables. In low-cost regions, seniors gain about 0.367 new discoveries, whereas juniors gain about 0.177, about 32% of the market. In medium-cost regions, although the overall increase is lower (in line with the model), the junior market share is higher, at about 56%. Therefore, the value of Landsat information to junior firms increases as their cost disadvantage increases. However, once cost differences increase substantially, the junior advantage goes down again. Figure 3(b) plots these market shares visually to validate that the empirical results fit the prediction that the Landsat maps help junior firms overcome cost

Table 4. Impact of Landsat on Juniors and Seniors by Cost Differences

	Post mapped		Post low-cloud	
Variables	Senior	Junior	Senior	Junior
Post × low cost	0.357***	0.152***	0.367***	0.177***
	(0.0562)	(0.0280)	(0.0578)	(0.0311)
Post × medium cost	0.148*	0.156***	0.138*	0.170***
	(0.0768)	(0.0517)	(0.0764)	(0.0522)
Post × high cost	0.0582**	−0.0156**	0.0477*	−0.000651
	(0.0297)	(0.00729)	(0.0270)	(0.00630)
Block fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	389,213	389,213	389,213	389,213

Notes. Standard errors clustered at block-level shown in parentheses. Block-year level observations, OLS models. In each of the three sets of specifications, I evaluate the impact of Landsat on Senior and Junior firms separately for regions with high-, medium-, and low-cost differences between junior and senior firms. These categories are created based on institutional quality rank at the block-level as described in Section 4.4.

$+p < 0.15$; $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

differences when those differences are low or medium, but these gains disappear in regions with very high costs. Combined with the results in the previous section, the data support the conclusion that not only does public data infrastructure benefit private-sector productivity as a whole, but it might be particularly valuable for entrant firms who have higher costs of investing in risky projects.

In addition to differences in cost, junior and senior firms could also differ in the extent to which they had access to private maps about their target regions. In the notation of the theory, seniors could have had a more informative prior, p_0, than juniors. In this case, one could imagine that the marginal value of Landsat map information would be higher for juniors, which could account for their increased market share. Although this mechanism is likely to be relevant for the impact of public data infrastructure generally, I do not find strong evidence for it in this context. Specifically, if we expected the differential information mechanism to hold, then we should expect that the impact of Landsat on junior market share should be the highest in places with little preexisting mapping, such as in countries with poor quality local institutions. In practice, I find an inverted U-shaped relationship between junior market share and institutional quality. Furthermore, as shown in online appendix Figure D.5, the positive effects of Landsat in Africa and Asia, where I would expect the priors for junior firms to be particularly poor, are minimal. Although certainly not conclusive evidence, these results provide confidence in the theoretical model that focuses on cost differences as an important mechanism through which juniors might benefit disproportionately from public mapping infrastructure. Having said that, I am unable to directly measure the extent to which juniors and seniors were privately informed about different regions before the arrival of Landsat. Such a measure would offer a more direct test of this mechanism, which could operate in parallel with the cost mechanism I outline.

5. Discussion

> 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

Although the public sector has engaged in mapping and other forms of public data provision for centuries, this paper provides some of the first empirical evidence on the role of these investments in shaping private sector performance and competition. This work has implications for the literature on the role of public resources on private performance, which thus far has examined the impact of public finance and public technology. This study provides a theoretical and empirical framework for the analysis of public data on the private sector. In particular, beyond the direct effects of public information in shaping the efficiency of private investments, I show that public data could have an important second-order effect by stimulating entry. Public data offers a relatively cheap instrument for the public sector to shape private performance and stimulate competition.

Furthermore, the process of innovation is often thought of as a search for solutions in complex landscapes (Fleming and Sorenson 2004, Aharonson and Schilling 2016)—not unlike the search for gold—and the language of exploration and discovery is common in the literature (March 1991, Henderson and Cockburn 1996, Manso 2011). By highlighting public data as a channel through which public investment shapes private discovery, my results add to previous work in this area that has highlighted the role of other channels such as IP restrictions, libraries, information technology, and financial incentives (David et al. 2000, Furman and Stern 2011, Kleis et al. 2012, Nagaraj 2017, Furman et al. 2018).

This work also has practical implications for managers, governments, and nonprofits looking to use data to shape private performance. The trove of government information (such as weather data, environmental information, consumer surveys, etc.) is large, as are other sources of free data such as Wikipedia, OpenStreetMap, and so on (Nagaraj 2020). Founders and startups in particular could look to exploit underused public domain information as an explicit strategy to reduce risk in the early stages of their project, thereby reducing the competitive advantage of larger firms. Similarly, governments and funders looking to foster private performance and entry might consider supporting large-scale mapping efforts that provide new public data on biological or physical entities (e.g., the Human Cell Atlas project maps all cells in the human body; Rozenblatt-Rosen et al. 2017, Nagaraj and Stern 2020). This study provides a theoretical framework and some of the first empirical evidence of the value of such efforts for downstream investment.

Finally, despite the contributions outlined previously, a few limitations of this study must be acknowledged. First, it is important to note that the Landsat variation used in the study is limited to about 25% of the earth's surface and that in the gold exploration industry defining a discovery can be somewhat arbitrary and hard to define. The results do seem robust to a battery of tests that investigate this issue. Second, although I focus on the role of costs as the primary mechanism to explain the heterogeneous effects of public data, future research should look at complementary channels as well. In particular, I do not investigate the likely possibility that public data could help smaller firms more than larger firms because smaller firms have less private information to begin with, because they lack technological or human capital or because they are better able to draw on academic science inspired by Landsat (Rowan 1975, Stuart and Podolny 1996). Finally, because I do not possess data on costs of extraction, one cannot conclude from the market share result that juniors benefit more from Landsat in terms of profitability or social welfare. Follow-up work should perform a more thorough examination of the welfare impacts of public data infrastructure, including cost and generalizability considerations across a variety of different programs.

Ultimately, this research suggests that public data offers a powerful, relatively cheap, and effective channel for governments to encourage private investment and performance and for new entrants to gain competitive advantage in dynamic markets. A more thorough understanding of this potentially important channel remains an exciting avenue for future work.

Acknowledgments

The author thanks Scott Stern, Heidi Williams, Catherine Tucker, and Pierre Azoulay for support and encouragement; Gustavo Manso, Ganesh Iyer, Toby Stuart, Sameer Srivastava, Aruna Ranganathan, Mathijs de Vaan, Michael Bikard, Joshua Krieger, Shane Greenstein, Jeff Furman, Dan Fehder, and Lisa McKay for comments; many other colleagues and seminar participants at the National Bureau of Economic Research Summer Institute, Massachusetts Institute of Technology Sloan Working Group, and the Massachusetts Institute of Technology Applied Economics Seminar for helpful comments; and several individuals from the National Aeronautics and Space Administration and the United States Geological Survey Landsat program and the gold exploration and remote-sensing industry, including Richard Schodde, for providing guidance and help with data. Petr Martynov and Melissa Staha provided excellent research assistance. All errors remain the author's own.

Endnotes

[1] The full talk can be accessed at https://www.youtube.com/watch?v=of8ai1HhqK4.

[2] My interview with an EROS center employee suggests that the data on the use of these images by firms were highly sensitive and have since been destroyed (personal communication, March 24, 2015). As such, it is unavailable for use in this research.

[3] A commercial satellite imagery provider was launched in the late 1980s through the *Satellite Pour Observation de la Terre* ("SPOT") satellite system (Chevrel et al. 1981).

[4] See http://landsat.usgs.gov/metadatalist.php.

References

Agrawal A, Galasso A, Oettl A (2017) Roads and innovation. *Rev. Econom. Statist.* 99(3):417–434.

Agrawal A, Rosell C, Simcoe TS (2014) *Do Tax Credits Affect R&D Expenditures by Small Firms? Evidence from Canada* (National Bureau of Economic Research, Cambridge, MA).

Aharonson BS, Schilling MA (2016) Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Res. Policy* 45(1): 81–96.

Arora A, Cohen WM (2015) Public support for technical advance: the role of firm size. *Industry Corporate Change* 24(4):791–802.

Arora A, Belenzon S, Patacconi A (2015) Killing the golden goose? The decline of science in corporate R&D. Technical report, National Bureau of Economic Research, Cambridge, MA.

Arora A, Belenzon S, Cohen W, Patacconi A (2019) Companies persist with biomedical papers. *Nature* 569(7756):S18–S19.

Ashley RP, Goetz AFH, Rowan LC, Abrams MJ (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, Menlo Park, CA.

Azoulay P, Zivin JSG, Li D, Sampat BN (2015) Public R&D investments and private-sector patenting: Evidence from NIH funding rules. Technical report, National Bureau of Economic Research, Cambridge, MA.

Berry S, Reiss P (2007) *Empirical Models of Entry and Market Structure. Handbook of Industrial Organization 3, 1845–1886* (Elsevier, New York).

Bikard M, Marx M (2020) Bridging academia and industry: How geographic hubs connect university science and corporate technology. *Management Sci.* 66(8):3425–3443.

Bresnahan TF, Reiss PC (1990) Entry in monopoly market. *Rev. Econom. Stud.* 57(4):531–553.

Brown WO, Burdekin RCK (2000) Fraud and financial markets: the 1997 collapse of the junior mining stocks. *J. Econom. Bus.* 52(3): 277–288.

Brynjolfsson E, McElheran K (2016) The rapid adoption of data-driven decision-making. *Amer. Econom. Rev.* 106(5):133–139.

Brynjolfsson E, Hitt LM, Kim HH (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? Working paper, Massachusetts Institute of Technology, Boston.

Card D, Chetty R, Feldstein MS, Saez E (2010) Expanding access to administrative data for research in the United States. NSF SBE 2020 White Paper, National Science Foundation Directorate of Social, Behavioral, and Economic Sciences, Arlington, VA.

Chevrel M, Courtois M, Weill G (1981) The SPOT satellite remote sensing mission. *Photogramming Engrg. Remote Sensing* 47: 1163–1171.

Christin A (2019) What data can do: A typology of mechanisms. *Internat. J. Comm.* 14:1115–1134.

Cohen WM, Nelson RR, Walsh JP (2002) Links and impacts: The influence of public research on industrial R&D. *Management Sci.* 48(1):1–23.

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

David PA, Hall BH, Toole AA (2000) Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Res. Policy* 29(4):497–529.

Draeger WC, Holm TM, Lauer DT, Thompson RJ (1997) The availability of Landsat data: Past, present, and future. *Photogramming Engrg. Remote Sensing* 63(7):869–875.

Ewens M, Nanda R, Rhodes-Kropf M (2018) Cost of experimentation and the evolution of venture capital. *J. Financial Econom.* 128(3): 422–442.

Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8-9):909–928.

Fleming L, Greene H, Li G, Marx M, Yao D (2019) Government-funded research increasingly fuels innovation. *Science* 364(6446): 1139–1141.

Forman C, Goldfarb A, Greenstein S (2012) The Internet and local wages: A puzzle. *Amer. Econom. Rev.* 102(1):556–575.

Furman JL, Stern S (2011) Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *Amer. Econom. Rev.* 101(5):1933–1963.

Furman JL, Nagler M, Watzinger M (2018) Disclosure and subsequent innovation: Evidence from the patent depository library program. Technical report, National Bureau of Economic Research, Cambridge, MA.

Gandhi SM, Sarkar BC (2016) Geological exploration. Gandhi SM, Sarkar BC, eds. *Essentials of Mineral Exploration and Evaluation* (Elsevier, New York),159–198.

Goward S, Arvidson T, Williams D, Faundeen J, Irons J, Franks S (2006) Historical record of Landsat global coverage. *Photogramming Engrg. Remote Sensing* 72(10):1155–1169.

Graham SJ, Hancock G, Marco AC, Myers AF (2013) The USPTO trademark case files data set: Descriptions, lessons, and insights. *J. Econom. Management Strategy* 22(4):669–705.

Hart M (2014) *Gold: The Race for the World's Most Seductive Metal* (Simon and Schuster, New York).

Henderson R, Cockburn I (1996) Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *RAND J. Econom.* 27(1):32–59.

Hill R, Stein C (2019) *Scooped! Estimating Rewards for Priority in Science* Working paper, Massachusetts Institute of Technology, Boston.

Howell ST (2017) Financing innovation: Evidence from R&D grants. *Amer. Econom. Rev.* 107(4):1136–1164.

Humphreys D (2016) *The Remaking of the Mining Industry*, 1st ed. (Palgrave Macmillan, Hampshire, UK).

Jackson T (2014) Fraser Institute annual survey of mining companies. Report, Fraser Institute, Vancouver, Canada.

Jayaraj S, Gittelman M (2018) Scientific Maps and Innovation: Impact of the Human Genome on Drug Discovery Working paper, Rutgers Business School, New Brunswick, NJ.

Kao J (2019) Charted Territory: Evidence from Mapping the Cancer Genome and R&D Decisions in the Pharmaceutical Industry.

Kleis L, Chwelos P, Ramirez RV, Cockburn I (2012) Information technology and intangible output: The impact of IT investment on innovation productivity. *Inform. Systems Res.* 23(1):42–59.

Krohn MD, Abrams MJ, Rowan LC (1978). Discrimination of hydrothermally altered rocks along the Battle Mountain-Eureka, Nevada, mineral belt using Landsat images. Technical report, US Geological Survey, Washington, DC.

Lerner J (2000) The government as venture capitalist: The long-run impact of the SBIR program. *J. Private Equity* 3(2):55–78.

Mack PE (1990) *Viewing the Earth: The Social Construction of the Landsat Satellite System* (MIT Press, Cambridge, MA).

Mansfield E (1986) The R&D tax credit and other technology policy issues. *Amer. Econom. Rev.* 76(2):190–194.

Manso G (2011) Motivating innovation. *J. Finance* 66(5):1823–1860.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Marco AC, Myers A, Graham SJ, D'Agostino P, Apple K (2015). The USPTO patent assignment data set: Descriptions and analysis. Working paper, United States Patent and Trademark Office, Alexandria, VA.

McCahon K, Fredricksen L (2014) *Fraser Institute Annual Survey of Mining Companies* (Fraser Institute).

Micek B (2017) How Entrepreneurs are Using Open Data to Start Businesses. Carto Blog. Accessed December 28, 2020, https://carto.com/blog/how-entrepreneurs-using-open-data-start-businesses/.

Mowery DC, Rosenberg N (1999) *Paths of Innovation: Technological Change in 20th-Century America* (Cambridge University Press, Cambridge, UK).

Munnell AH (1992) Policy watch: Infrastructure investment and economic growth. *J. Econom. Perspective* 6(4):189–198.

Nagaraj A (2017) Does copyright affect reuse? Evidence from Google Books and Wikipedia. *Management Sci.* 64(7):2973–3468.

Nagaraj A (2020) Information seeding and knowledge production in online communities: Evidence from openstreetmap. *Management Sci.* Forthcoming.

Nagaraj A, Stern S (2020) The economics of maps. *J. Econom. Perspective* 34(1):196–221.

Nagaraj A, De Vaan M, Shears E (2020) Improving data access democratizes science. *Proc. Natl. Acad. Sci. USA* 117(38):23490–23498.

Nelson RR (1982) The role of knowledge in R&D efficiency. *Quart. J. Econom.* 97(3):453–470.

Ray WA (1997) Policy and program analysis using administrative databases. *Ann. Internal Medicine* 127(8-2):712–718.

Rowan LC (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. *Amer. Sci.* 63(4):393–403.

Rowan LC, Goetz AF, Ashley RP (1977) Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images. *Geophysics* 42(3):522–535.

Rozenblatt-Rosen O, Stubbington MJ, Regev A, Teichmann SA (2017) The Human Cell Atlas: From vision to reality. *NATNEWS* 550(7677):451.

Schodde R (2011) Recent trends in gold discovery. Presentation, NewGenGold Conference, November 23, Paydirt Media, Perth, AU.

Schodde R (2015) Canada's discovery performance and outlook. Presentation, Prospectors & Developers Association of Canada (PDAC) Meetings, March 3, Toronto, ON.

Scotchmer S (2004) *Innovation and Incentives* (MIT Press, Cambridge, MA).

Seamans R, Sohn E, Sands D (2017) Technological opportunity and the locus of innovation: Airmail, aircraft, and local capabilities. Working paper, New York University Stern School of Business, New York.

Slade ME (2001) Valuing managerial flexibility: An application of real-option theory to mining investments. *J. Environ. Econom. Management* 41(2):193–233.

Spurr SH (1954) History of forest photogrammetry and aerial mapping. *Photogramming Engrg.* 20:551–560.

Stuart TE, Podolny JM (1996) Local search and the evolution of technological capabilities. *Strategic Management J.* 17(S1): 21–38.

Ubaldi B (2013) Open government data: Toward empirical analysis of open government data initiatives. Working paper, Organisation for Economic Co-Operation and Development, Paris.

Verhulst S, Young A (2016) The Global Impact of Open Data (O'Reilly Media, Sebastopol, CA).

Vincent RK (1975) Commercial applications of geological remote sensing. Cruz JB, Pearson JB, Haddad AH, eds. *Proc. Decision and Control Including the 14th Sympos. on Adapt. Processes*, 1975 (Institute of Electrical and Electronics Engineers, Washington, DC), 258–263.