The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXI, NO. 1 • FEBRUARY 2016

Can Markets Discipline Government Agencies? Evidence from the Weather Derivatives Market

AMIYATOSH PURNANANDAM and DANIEL WEAGLEY*

ABSTRACT

We analyze the role of financial markets in shaping the incentives of government agencies using a unique empirical setting: the weather derivatives market. We show that the introduction of weather derivative contracts on the Chicago Mercantile Exchange (CME) improves the accuracy of temperature measurement by 13% to 20% at the underlying weather stations. We argue that temperature-based financial markets generate additional scrutiny of the temperature data measured by the National Weather Service, which motivates the agency to minimize measurement errors. Our results have broader implications: the visibility and scrutiny generated by financial markets can potentially improve the efficiency of government agencies.

HOW ARE THE INCENTIVES OF GOVERNMENT agencies shaped? This question is of fundamental importance to both economists and policy makers. In this paper, we examine a novel channel through which government agencies' behavior may be affected: financial market pressure. Specifically, we ask whether government bureaucrats improve their performance when their actions are closely scrutinized because they affect financial markets? We address this broad question by exploiting an interesting empirical setting: the launch of a new financial market, namely, exchange-traded weather derivatives, that has payoffs linked to the measurement of temperature by a government agency—the National Weather Service (NWS). Financial markets depend on the actions of government agencies in a number of other settings as well, such as commodity futures and catastrophic insurance markets.[1] We focus on the weather derivatives

*Purnanandam is at the Ross School of Business, University of Michigan. Weagley is at the Scheller College of Business, Georgia Institute of Technology. The authors thank Gautam Ahuja; Taylor Begley; Sugato Bhattacharyya; Ing-Haw Cheng; Charlie Hadlock; Zoran Ivkovic; Vojislav Maksimovic; Paolo Pasquariello; Uday Rajan; Michael Roberts; Tyler Shumway; René Stulz; Maciej Szefler; Vish Viswanathan; two anonymous referees; an anonymous associate editor; and seminar participants at Dartmouth, Michigan State University, University of Michigan, and NBER meetings on Economics of Commodities Markets for helpful comments. We thank CME and MDA Inc. for providing data and several clarifications on the weather derivatives market. The authors are responsible for all remaining errors. We have received no external financial support for this research and are awave of conflicts of interest.

[1] For example, a number of dairy, livestock, and commodity contracts on the CME and Chicago Board of Trade (CBOT) are settled based on measures provided by the U.S. Department of Agriculture (USDA). Information provided by the U.S. Geological Survey (USGS) plays a crucial role in the pricing of earthquake insurance. Similarly, market participants rely on the National Oceanic

DOI: 10.1111/jofi.12366

market because it provides us with a clean empirical setting to address the central question.

The Chicago Mercantile Exchange introduced the first exchange-traded weather derivative instruments in 1999. Since then, the CME has introduced weather contracts on a number of U.S. cities in a staggered fashion. A vast majority of these instruments are temperature-related, allowing end-users to hedge their exposure to undesirable warm or cold weather conditions. These contracts are city-specific and are settled based on the temperature readings of a specific NWS weather station within or near the contract city. The weather stations are prone to measurement error due to factors such as improper calibration of the sensors, poor maintenance, and lax monitoring of the equipment. We examine the effect of CME weather derivative introduction on NWS temperature measurement performance. The introduction of derivative contracts directly ties the NWS-reported temperature at these stations to the large economic interests of traders and hedgers in the market. The resulting increase in market scrutiny can impose significant reputation losses on the NWS and its managers if the measurement is inaccurate. Thus, the introduction of a weather derivatives market provides a reasonably exogenous shock to market-based pressure on the NWS employees responsible for measuring the temperature at a designated weather station in the contract city.

As of June 30, 2012, there are 24 U.S. cities with temperature-related derivative contracts traded on the CME. These contracts were issued in four waves: 1999 to 2000, 2003, 2005, and 2008. Our empirical setting allows us to compare the improvement in temperature accuracy of the weather stations with derivatives (the treatment group) around the derivative launch dates with a set of nonderivative stations (the control group) during the same period. The staggered nature of the derivative launch allows us to separate improvements due to derivative introduction from the effect of any time trend in error rate or any general improvement in the NWS's technology over time. Furthermore, unlike stocks or bonds, the variable underlying the weather derivatives contract is not a traded commodity. Thus, the introduction of the derivative contract is not going to affect the underlying weather itself, which helps us establish a causal relationship from financial market pressure to measurement accuracy.

Our main performance measure captures temperature measurement errors made by the NWS. We define measurement error as any discrepancy between the initial temperature recording of the NWS weather station and the subsequent corrected values issued by the National Climatic Data Center (NCDC, a sister agency of the NWS) or a private third party. Using a sample period of 1999 to 2012 for 49 treatment and control stations, we find that the median weather station in our sample has an error rate of 12 days per year. Using a difference-in-differences model, we show that the treated station's error rate

and Atmospheric Administration (NOAA) for assessing, pricing, and settling a number of contracts such as hurricane and flood insurance. At a much broader level, across the world's markets, participants focus closely on government-released macroeconomic data such as GDP growth, inflation, and unemployment for their investment strategies.

declines significantly after the introduction of weather derivative contracts. Depending on the model specification, the point estimate ranges from 1.6 to 2.4 days of improvement in measurement accuracy at the treated stations. The decline in error rate represents about 13% to 20% of the median error rate in our sample. Our results remain similar if we omit the nonshocked stations from the sample, and thus achieve identification solely from the group of derivative stations. On a station-by-station basis, we find that 18 of 20 stations experience an improvement in error rate after weather derivative introduction. Furthermore, we show that our results are not driven by differential trends across the treatment and control groups. Taken together, these results paint a clear picture: weather stations with derivative contracts have a lower incidence of inaccurate data after their recorded temperature numbers become reference points for billions of dollars of financial contracts in an open market.

In our next set of tests, we show that the improvement is larger for stations with higher economic interests. Specifically, we find the improvement to be higher for stations that received derivative contracts in earlier waves. These stations are likely to have relatively higher economic interests based on CME's revealed preferences. Moreover, we find that the effects are stronger for cities that are likely to have higher energy demand and hence higher interest in weather derivative products.

Our results are consistent with the idea that bureaucracies are concerned about a loss in reputation and improve performance when the probability of a loss in reputation increases. The additional visibility and scrutiny of an agency's actions by market forces can create significant reputational risks for the agency, which in turn can have adverse consequences for the agency's budget and its managers' career paths (Wilson (1989)).[2] Consequently, these concerns are likely to motivate the bureaucrats to control managerial slack in response to increased pressure from financial markets.[3]

An alternative channel that can explain our main result is based on technological improvements at the treatment stations. If the NWS deploys more technical resources at the treatment stations precisely at the time of derivative launch, then we are likely to observe lower error rates even without any improvement in managerial effort. While we cannot conclusively establish the effect of the managerial effort channel, we provide several tests to show that the technological channel is unlikely to explain all of our results. First, we show

[2] See Dewatripont, Jewitt, and Tirole (1999) for a discussion on the importance of career concerns in bureaucracies.

[3] This mechanism is different from the traditional models of corporate finance that study the disciplining role of financial markets in the context of corporate managers. Numerous important contributions in this area have been nicely summarized in survey articles such as Shleifer and Vishny (1997), Gillan and Starks (1998), Black (1998), Karpoff (1998), Romano (2001), Hermalin and Weisbach (2003), and Becht, Bolton, and Röell (2003), among others. This line of research argues that market participants such as blockholders and pension funds can discipline corporate managers through explicit or implicit performance-based incentive contracts. See Shleifer and Vishny (1986), Holmström and Tirole (1993), Burkart, Gromb, and Panunzi (1997), Bolton and Von Thadden (1998), Kahn and Winton (1998), Gopalan (2009), Admati and Pfleiderer (2009), and Edmans (2009), among others.

that incremental technological investments made by the NWS do not correlate with the derivative launch dates. This shows that the treated weather stations achieve more accurate measurement with similar resources as the control stations. In another test, we examine measurement improvement across trading months with high and low trading activity. If the NWS selectively introduces better equipment at the treated stations at the time of derivative launch, then the improvement in measurement accuracy should be observed throughout the year. Instead, we find that all of the accuracy improvements come precisely from the peak trading months, when market scrutiny is high. The results broadly support the managerial effort channel.

We assess the economic importance of these effects by computing how the improvement in measurement accuracy affects the difference in the weather derivative contract payoffs if the contracts were to settle on the raw versus corrected temperature numbers. We estimate the effect to be in the range of 21 to 65 basis points of the notional value of a representative contract. While these numbers are small in the context of the entire financial market, they are likely to be meaningful for participants in this market, especially for financial speculators that operate with very high leverage and invest in the market based on a margin requirement of 5% to 10% of the contract's value. The improvement in temperature measurement can have a significant positive effect on hedgers in this market as well. The more reliable the reference point for contract settlement, the more willing hedgers will be to participate in the market, as it decreases basis risk. Consistent with this idea, we find that weather stations with lower error rates have higher contract volume in the market. These findings show that the effects are economically meaningful for market participants, and hence the participants are likely to focus on the accuracy of temperature recordings.

While we focus on a specific market for drawing sharp empirical conclusions, our study has possible implications for both the efficiency of public enterprise more generally (see Karpoff (2001)) and financial markets' role in making public bureaucracies more efficient. Our study has limitations, however, in terms of its external validity. Specifically, our study is narrow relative to other government actions that can be affected by financial markets, such as the formulation of macroeconomic policies and measurement of macroeconomic indicators. We leave the study of additional interactions between government agencies and financial markets for future research.

Our study also contributes to our understanding of the role of financial innovation and derivatives in affecting real outcomes.[4] We do not carry out a

[4] See Tufano (2003) for a survey on financial innovation, including the role of innovation on society. See Stulz (2004) for a survey of the literature and discussions on the costs and benefits of derivatives. Pérez-González and Yun (2013) show that weather risk-management derivative usage leads to higher valuation, investment, and leverage for energy utilities. There is a large literature on the effect of financial derivatives on firm valuation and investment decisions in nonweather risk-related contexts as well. For example, see Allayannis and Weston (2001), Carter, Rogers, and Simkins (2006), Purnanandam (2007), and Berrospide, Purnanandam, and Rajan (2010), among others.

detailed analysis of all the costs and benefits of improvements in temperature measurement, and thus we are unable to make any welfare claims. However, if the improved accuracy comes at little or no incremental costs, there is a potential for welfare gains due to better risk-sharing in the economy. Pérez-González and Yun (2013) directly analyze the beneficial effects of weather derivative contracts for energy utilities. Their study shows that utilities can lower their cost of capital and bankruptcy risk by using these contracts. If the improvement in measurement accuracy improves the viability of weather derivative markets, then such hedgers can achieve better risk-sharing with the rest of the economy.

The rest of the paper is organized as follows. Section I elaborates on the economic channel of interest. Section II describes the weather derivatives market in detail and highlights key aspects of temperature measurement by the NWS. Section III describes the data and provides sample statistics. Section IV discusses the empirical design and presents results. Section V concludes.

I. Economic Channel

The goal of our paper is to study the role of financial markets in shaping the incentives of government agencies. If the payoffs of certain financial contracts depend on the actions of a government agency, then market participants are likely to pay close attention to the value-relevant activities of the agency. Prendergast (2007) notes that the clients of a bureaucracy typically point out errors when it harms them. For financial contracts, especially for zero-sum derivative contracts, it is likely that some investors stand to lose from inaccurate reports made by the agency and, hence, they are motivated to point out the agency's mistakes.

Increased scrutiny and visibility of an agency's actions are likely to increase its concerns about reputational losses, with consequences such as political hearings, budget cuts, and downsizing (Wilson (1989)). In addition, reputation loss can have an adverse impact on the career of the bureaucracy's managers. Furthermore, if the agency's actions have implications for the payoffs in financial markets, there is a higher probability of disputes among market participants in the case of any inaccuracy in the agency's actions—while the government agency may not be a party to any resulting litigation, it may experience negative publicity or a loss of reputation due to the lawsuit. All these concerns should motivate the bureaucracy to control managerial slack in its agencies in response to an increase in scrutiny by financial markets.

In the weather derivatives market, accurate temperature measurement is likely important to all market participants: investors, hedgers, and the trading exchange (CME). There are some key features of the weather derivatives market that create strong incentives for market participants to pay close attention to the agency's actions. First, due to the zero-sum nature of derivative payoffs, it is likely that at least some market participants are always motivated to point out measurement errors by the NWS. Second, several participants in this market are highly levered financial institutions. The leverage effect can significantly amplify small differences in contract payoffs. And, finally, gains and losses in the derivatives market due to relatively inaccurate temperature data are relatively straightforward to compute, as compared to other settings such as potential losses in agriculture output.[5] These factors should motivate financial market participants to monitor NWS actions and point out inaccuracies. In summary, the increase in scrutiny and dollar values tied to weather derivative contracts after introduction increases the reputational concerns of the NWS, which should lead to improved temperature measurement by the agency.

It is worth emphasizing that there is a fundamental difference between our setting, which involves a government agency, and the traditional role of the market in disciplining corporate managers through compensation contracts.[6] A canonical model in the corporate literature, such as Holmström and Tirole (1993), assumes that the financial markets receive an informative signal about a manager's performance. As market participants trade based on this signal, the value-relevant information gets incorporated into the firm's stock price. The manager's compensation, in turn, is linked to the stock price. Thus, the manager stands to lose if he makes a non-value-maximizing decision. In the corporate setting, the agent is disciplined by the market because a bad decision will lower the stock price and the agent will suffer through decreased compensation.[7] Government agencies do not have stock prices and hence there are no stock-based compensation contracts for their managers. Instead, the economic mechanism behind our study is based on the reputational concerns of the bureaucracy.

II. Weather Derivatives Market

Weather has a significant impact on the operating and financial performance of several industries, municipalities, and households. Some survey evidence suggests that over 30% of U.S. GDP is associated with weather-sensitive industries (Dutton (2002)). While the need for insurance against weather conditions has existed for a long time, the first set of exchange-traded weather contracts was listed on the CME in 1999. The exchange launched temperature-based futures and options contracts on 10 U.S. cities within 13 months of

[5] It is possible that there are other non-market-based settings, like agriculture production, where the stakes on temperature measurement accuracy increase for reasons unrelated to financial payoffs. However, it is unlikely that the stakes increase at the same time as derivative introduction, considering the staggered nature of weather derivative contract introduction.

[6] Furthermore, the absence of a corporate control market and a relative lack of competitive forces makes public bureaucracies different from their private sector counterparts. Researchers have argued that the lack of these mechanisms makes them more inefficient as compared to their private counterparts (Boardman and Vining (1989)). The lack of these mechanisms has been at the center of a long-standing debate in the performance evaluation of public bureaucracies (see Heckman, Heinrich, and Smith (1997) for further discussions).

[7] It has long been argued that financial markets can influence real decisions through channels such as information aggregation (Hayek (1945)) and optimal risk-sharing across agents in the economy (Allen and Gale (1994)). Our focus is on the effect of financial markets in minimizing managerial slack.

Figure 1. Derivative introduction dates. This figure shows the introduction years of U.S. weather derivatives listed on the CME. For each year, we list the locations that received a derivative.

September 1999. It subsequently launched contracts on several other cities in three more waves in 2003, 2005, and 2008. As of June 30, 2012, CME weather contracts are available for 24 U.S. cities spanning all broad meteorological areas of the country.[8] We provide a timeline of the introduction of these contracts in Figure 1.

An overwhelming majority of weather contracts are based on temperature.[9] Temperature-related contracts insure the buyers from either excessive heat or excessive cold during a specified period of time. The two main temperature contracts are Heating Degree Day (HDD) and Cooling Degree Day (CDD) contracts. The buyer of an HDD contract receives payments for cold days, defined as days with average temperature below 65°F; conversely, the buyer of a CDD contract receives payments for hot days, defined as days with average temperature exceeding 65°F. These contracts are written on the observed temperature at a specific weather station near the contract city for a specific period. The period can be a month, week, or season. In 2005, approximately the middle point of our sample period, the total outstanding notional value of all CME-traded monthly weather contracts amounted to about $22 billion. The number increases to approximately $43 billion if we include seasonal and other nonmonthly contracts as well.

An Example Contract: As an illustration, consider a CDD contract on Chicago for the month of August. The contract settles on the temperature number

[8] In addition to these 24 cities, CME has snowfall contracts on Newark and the hurricane index on the eastern United States from Brownsville, Texas to Eastport, Maine.

[9] Based on survey evidence, the Weather Risk Management Association (WRMA) reported that over 95% of the CME contracts, in notional value terms, were related to temperature in 2005 to 2006 (WRMA Survey Report, PricewaterhouseCoopers (2006)). Other major categories included contracts on rain, wind, and snow.

reported by the weather station located at O'Hare International Airport, with its unique Weather Bureau Army Navy (i.e., station, WBAN) number 94846.[10] For every day in August, the CDD contract compares the average of the daily maximum and minimum temperatures (T_{avg}) reported at this station to 65°F and computes the cooling degrees for the day as max[0, $T_{avg} - 65$]. These degree days are cumulated over the entire month of August and payments are made based on the cumulative month-end number, called the CDD index, for August. Typically, one point in the index entitles the buyer to a payment of $20 from the seller. With hundreds of thousands of such contracts in the market, the reported temperature at these stations has considerable economic implications for the market participants.

The final settlement of contracts is based on the CDD or HDD index reported by MDA Information Systems, Inc. Settlement occurs on the second business day after the contract month.[11] MDA (formerly Earth Satellite Corporation, founded in 1969), a private company and leading provider of weather data to the weather trading industry, obtains weather data generated by the NWS at each weather station and performs several quality control checks before transmitting temperatures to the CME for trade settlements. MDA's quality checks are based on cross-verification, consistency of the data with other nearby stations, and their own meteorological models.

A. Temperature Measurement and Sources of Error

The NOAA is the main government bureau responsible for the measurement and monitoring of surface and ocean weather in the country. The NWS, an agency within NOAA, is responsible for monitoring surface temperatures, including the production and dissemination of temperature readings. The NCDC, another agency within NOAA, archives and processes past weather records.

NOAA and the NWS have detailed procedure manuals for collecting temperature readings in a timely and accurate manner.[12] They lay out procedures for proper installation, monitoring, and maintenance of weather stations. A few examples of these guidelines are: (i) the instrument must be placed at least 100 feet from any concrete or paved surface; (ii) all attempts should be made to avoid areas with rough terrain, air drainage, areas where water tends to collect, and areas where drifting snow collects; and (iii) the instrument should not have any major obstruction (for example, nearby buildings, trees, or fence) close by that can affect its readings. These instructions point out the possible

[10] WBAN, an acronym for Weather Bureau Army Navy, is a five-digit weather station number that uniquely identifies a measurement location.

[11] See the guidelines on CME's website at: http://www.cmegroup.com/trading/weather/files/Monthly-CDD-Index-Futures-Final-Settlement-Procedure.pdf.

[12] They also issue regular directives to their field offices on best practices in measuring temperature. These directives can be obtained from NOAA's website at: http://www.nws.noaa.gov/directives/010/010.htm. As an example, consider NWS instruction 10-1302, dated June 21, 2010. It details the requirements and standards for NWS temperature and precipitation recordings. See http://www.nws.noaa.gov/directives/sym/pd01013002curr.pdf.

sources of error in temperature measurement and the NWS's attempts to train its staff to minimize these error rates. We provide an example of measurement error at the Chicago O'Hare station in the Internet Appendix of the paper.[13] In the rest of the paper, we empirically analyze the effect of weather derivative launch on the frequency of such errors at the underlying stations.

III. Data and Descriptive Statistics

We obtain information on the launch dates of monthly derivative contracts on a city's temperature from the CME and press releases.[14] The list of derivative stations covers highly populated cities as well as a few smaller cities that are likely to have large economic interests tied to weather. We identify 25 stations without weather derivative contracts as the control group. The 25 control weather stations are chosen by sorting all U.S. metropolitan areas by population and using the 25 highest population cities without weather derivatives.[15] We identify the WBAN number of all derivative cities based on the contract specification. For the control cities, we use the weather station at the largest nearby airport since 23 of 24 CME derivative stations are located at the contract city's main airport. In total, we have 49 weather stations in our sample. These weather stations, their WBAN identification number, and the derivative introduction dates for the treatment group are provided in Table I.

We obtain weather data for these stations from MDA Information Systems. We obtain two pieces of information from MDA: (i) raw temperature readings, and (ii) "cleaned" or "corrected" temperature values. For both sets of data, we obtain daily maximum and minimum temperatures because the weather derivative contracts are settled based on the average of these two values. The raw temperature readings are the actual reported temperature numbers by the NWS or an affiliated organization for each station on a given day. The raw temperature comes from METAR readings, which are standardized weather reports produced by the Automated Surface Observing Systems (ASOS) located at the respective weather stations.[16] The cleaned or corrected temperature value for every station-date pair is produced by MDA using a detailed five-step

[13] The Internet Appendix may be found in the online version of this article.

[14] For some cities, the CME introduced weekly and seasonal contracts at a later date as well. These contracts were introduced after the monthly contracts, and thus we focus on the monthly contract introduction dates.

[15] We use the 2011 population estimates for metropolitan areas from the U.S. Census Bureau for this purpose.

[16] These stations are collectively operated by the Federal Aviation Administration, NWS, and the Department of Defense. For expositional simplicity, we call these stations NWS-operated stations throughout the paper since they are the main nodal agency for temperature-related activities. MDA obtains the raw temperature data for each weather station from the NWS METAR reports. The NWS stations produce hourly weather reports, six-hour min/max temperature reports, and 24-hour min/max temperature reports at midnight local time. We obtain the 24-hour min/max temperature values as the measures of raw temperature. If this value is not available for a specific station-date, then MDA provides us with the minimum and maximum temperature based on six-hour or hourly reports.

Table I

Weather Stations

This table presents information on the weather stations in our sample. Our sample includes 24 derivative stations (Panel A) and 25 control stations (Panel B). Station Location is the specific location of the weather station. WBAN is the Weather Bureau Army Navy station identifier. Intro. Date is the first trading day of a temperature-based weather derivative contract settling based on the station's readings. Pop. Rank is the 2011 U.S. population rank of the nearest metropolitan area. Service is the weather station's service level classification by the National Weather Service. Service levels range from A to D, with A being the highest service level. Traffic is the 2011 airport passenger traffic rank among U.S. airports.

Station Location	WBAN	Intro. Date	Pop. Rank	Service	Traffic
Panel A. Weather Derivative Stations					
Atlanta, GA	13874	Sep. 22, 1999	9	A	1
Chicago, IL	94846	Sep. 22, 1999	3	A	2
Cincinnati, OH	93814	Sep. 22, 1999	27	A	>46
New York City, NY	14732	Sep. 22, 1999	1	A	20
Dallas-Fort Worth, TX	3927	Sep. 30, 2000	4	A	4
Des Moines, IA	14933	Sep. 30, 2000	88	A	>46
Las Vegas, NV	23169	Sep. 30, 2000	30	A	9
Philadelphia, PA	13739	Sep. 30, 2000	6	A	18
Portland, OR	24229	Sep. 30, 2000	23	A	30
Tucson, AZ	23160	Sep. 30, 2000	52	C	>46
Boston, MA	14739	Sep. 26, 2003	10	A	19
Houston, TX	12960	Sep. 26, 2003	5	A	10
Kansas City, MO	3947	Sep. 26, 2003	29	A	32
Minneapolis-St. Paul, MN	14922	Sep. 26, 2003	16	A	15
Sacramento, CA	23232	Sep. 26, 2003	25	C	>46
Baltimore, MD	93721	June 20, 2005	20	A	23
Detroit, MI	94847	June 20, 2005	13	A	17
Salt Lake City, UT	24127	June 20, 2005	48	A	24
Colorado Springs, CO	93037	May 19, 2008	81	C	>46
Jacksonville, FL	13889	May 19, 2008	40	A	>46
Little Rock, AR	13963	May 19, 2008	72	A	>46
Los Angeles, CA	93134	May 19, 2008	2	–	3
Raleigh-Durham, NC	13722	May 19, 2008	47	A	38
Washington, DC	13743	May 19, 2008	7	A	25
Panel B. Control Stations					
Austin, TX	13904	–	34	A	39
Charlotte, NC	13881	–	33	A	11
Cleveland, OH	14820	–	28	A	37
Columbus, OH	14821	–	32	A	>46
Denver, CO	03017	–	21	A	5
Indianapolis, IN	93819	–	35	A	>46
Louisville, KY	93821	–	42	A	>46
Memphis, TN	13893	–	41	A	40
Miami, FL	12839	–	8	A	12
Milwaukee, WI	14839	–	39	A	35
Nashville, TN	13897	–	37	A	34
Norfolk, VA	13737	–	36	C	>46
Oklahoma City, OK	13967	–	43	A	>46
Orlando, FL	12815	–	26	A	13

(*Continued*)

Table I—*Continued*

Station Location	WBAN	Intro. Date	Pop. Rank	Service	Traffic
Panel B. Control Stations					
Phoenix, AZ	23183	–	14	A	8
Pittsburgh, PA	94823	–	22	A	45
Providence, RI	14765	–	38	A	>46
Riverside, CA	03171	–	12	C	>46
San Antonio, TX	12921	–	24	A	46
San Diego, CA	23188	–	17	B	28
San Francisco, CA	23234	–	11	A	7
San Jose, CA	23293	–	31	B	44
Seattle, WA	24233	–	15	A	16
St. Louis, MO	13994	–	19	A	31
Tampa, FL	12842	–	18	A	29

process to clean the raw temperature values obtained from the government agencies.[17] Through this process, MDA ensures that the data are consistent with nearby reporting stations. MDA also takes care of missing temperature values, which occur in the NWS reports due to reasons such as improper or incomplete METAR recordings. If the raw data have missing values, MDA uses other sources, such as NWS Climate Summary Reports, contacts at the local NWS office, or local media reports to obtain temperature values. MDA also checks the meteorological consistency of the reported data by checking the data against itself and against alternative data sources, such as surrounding stations. Using this detailed process, MDA arrives at a clean temperature measure that is widely used by the financial services industry as well as several other sectors. In essence, the MDA-cleaned values are third-party corrected temperature numbers for these weather stations.

We obtain the raw and cleaned data on a daily basis for all 49 stations in our sample over the 13-year period from 1999 to 2011. That the sample begins in 1999 should have only a minor impact on our study because 20 of the 24 treatment stations received derivative contracts after 1999, that is, we have data for both before and after derivative introduction for 20 stations.

We also obtain data on cleaned temperatures from the NCDC. The NCDC issues these official temperatures with a couple months' time lag. These corrections, or restatements, by the NCDC provide us with yet another measure of measurement accuracy at the time of initial report.[18] The NCDC restated numbers are extremely close to the MDA corrected values. We choose to use the difference between NWS raw numbers and MDA corrected values as the main variable in our tests. We prefer the MDA clean values because they alleviate

[17] Further details can be found in MDA's Procedure Manual: http://www.cmegroup.com/trading/weather/files/procedure-manual.pdf.

[18] Further information on preliminary and cleaned data can be obtained from NWS instruction manuals such as NWSI 10-1004 dated February 17, 2011 and NWSPS 10-10 dated September 29, 2010.

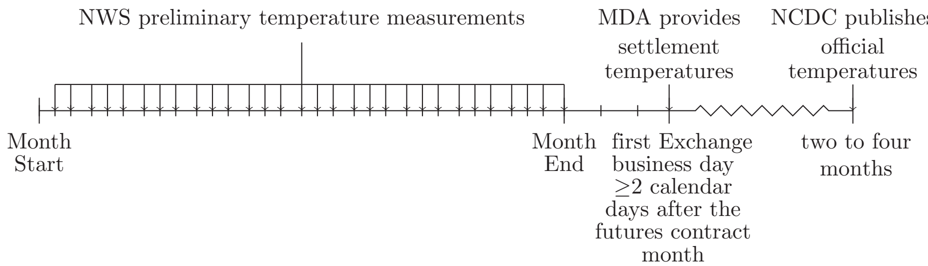

Figure 2. Weather measurement timeline. This figure shows the timeline of initial temperature measurement by the NWS, the corrections reported by MDA for contract settlement, and the final cleaned value generated by NCDC.

the concern that government agencies may be less inclined to restate their recordings after these contracts begin to trade. Figure 2 provides a timeline of initial temperature measurement by the NWS, the corrections reported by MDA for contract settlement by CME, and the final cleaned value generated by the NCDC.

A. Descriptive Statistics

We take the number of days in a year when a given station reports erroneous or missing values as the measure of temperature inaccuracy. These are the dates when the raw and corrected values differ from each other. Some station-years show a considerable error rate, resulting in skewness in the data. We winsorize the data at 5% from both tails to ensure that our results are not driven by outliers. We also use the log transformed error rate as an alternative dependent variable to alleviate concerns about outliers.

Summary statistics are presented in Table II. A representative median station reports about 12 error days per year. There is considerable cross-sectional variation in the data as evidenced by the 90$^{\text{th}}$ (20 days) and 10$^{\text{th}}$ (5 days) percentiles of error days in the sample. Raw and final numbers remain the same for 96.69% of days. Of the remaining 3.31% of days, 2.12% have a difference of 1°F between the raw and cleaned data. The remaining 1.19% have considerably large discrepancies, mostly ranging from 2°F to 10$^{\circ}F$.

B. Selection of Control Stations

We choose control stations based on the population of the city. A χ^2 test for the equality of population distribution across the treatment and control groups has a p-value of 0.43. In addition to their population, the treatment and control cities are comparable on several other relevant dimensions. Most important, they are similar in terms of the NWS's designated service level. The NWS classifies weather stations into four service levels (A, B, C, and D) based on air traffic and bad weather outcomes. Service Level D stations are completely

Table II

Summary Statistics for Total Weather Station Errors

This table presents summary statistics on weather station errors. Each observation is a weather station-year. The 49 stations consist of the 24 weather stations underlying a CME temperature contract and the 25 control weather stations. Year is the year for which the summary statistics are calculated. *N* is the number of weather stations in the sample during the year. *SD* is the standard deviation of the number of errors across stations during the year.

Year	*N*	Mean	Median	*SD*	10^{th}	90^{th}
1999	46	11.91	11.5	5.31	6	19
2000	49	11.49	11	5.40	4	19
2001	49	11.14	10	5.30	5	19
2002	49	15.51	16	5.21	8	23
2003	49	14.45	15	6.37	5	22
2004	49	13.57	13	5.83	6	23
2005	49	11.49	11	4.74	5	17
2006	49	10.57	9	4.82	4	17
2007	49	12.73	13	5.02	7	20
2008	49	12.47	12	4.59	5	19
2009	49	10.63	10	3.89	4	15
2010	49	10.49	10	4.36	4	16
2011	49	10.71	10	4.19	5	17
All	634	12.09	12	5.23	5	20

automated. Service Level C stations have an additional human observer when the tower is open. Service Levels A and B have a human observer practically all the time, and the observers have more responsibilities at these stations. As documented in Table I, almost all of our treatment and control stations belong to category A, that is, to the category that requires the utmost care from human observers.[19] A χ^2 test for the equality of distribution cannot reject the null hypothesis of equal distribution of service levels across the two groups (p-value=0.34).

As expected, these groups are comparable in terms of airport traffic as well.[20] The p-value of the χ^2 test for the equality of distribution is 0.47 for airport traffic. We also assess the comparability of historical temperature and other weather measures across the two groups. In the Internet Appendix, we document that the average as well as minimum and maximum temperatures across these two groups are statistically similar. The monthly precipitation, number of thunderstorm days, and number of fog days across the two groups are comparable and statistically indistinguishable. In sum, the control cities are similar to the treatment cities in terms of their designated service levels, nonderivative-related interests in temperature, and weather conditions. The control group is thus likely to serve as a reasonable counterfactual for our study.

[19] Service-level data come from the Aviation Weather Assets Database: http://apps.avmet.com/awad/AWADReport.cfm.

[20] Airport traffic ranks come from Airports Council International: http://www.aci-na.org/content/airport-traffic-reports.

IV. Empirical Design and Analysis

A. Research Design

We estimate the effect of weather derivative introduction on the accuracy of temperature measurement in a difference-in-differences framework. We compare the measurement accuracy of a weather derivative station after the shock (i.e., after the introduction date) to the same station's accuracy before the shock to get the first margin of difference. The second margin of difference comes from the change in the accuracy level of nonshocked stations around the same time period. The underlying assumption is that changes in the nonshocked stations' accuracy level separate out the effect of other (i.e., nonderivative-related) factors on the accuracy level of the shocked stations. These nonderivative-related changes in accuracy can come from sources such as technological advancement over time, climatic changes, or NWS's overall effort to improve its accuracy levels across all stations. A key advantage of the staggered launch of weather derivative contracts across cities is that it allows us to remove the effect of any such macroeconomic or broad climatic factors on measurement accuracy. We implement this research design using the following regression model with both station and year fixed effects:

$$y_{st} = \alpha_s + \beta \times Derivative_{st} + year_t + \epsilon_{st}, \tag{1}$$

where y_{st} denotes measurement error at NWS weather station s in year t, α_s stands for station fixed effects, and $year_t$ denotes year fixed effects. The independent variable $Derivative_{st}$ takes a value of one for station-year observations after the introduction of derivatives, and zero otherwise. The year of introduction is included in the postintroduction period. Station fixed effects remove the station-specific component of measurement error whereas year fixed effects control for broad time-specific effects, including the possibility of any secular improvement in measurement accuracy across all stations. Thus, the coefficient on $Derivative_{st}$ provides the difference-in-differences estimate of interest.

The key identifying assumption behind our empirical exercise is that the launch of a weather derivative is uncorrelated with any unobserved improvement in the station's ability to measure the temperature. It is highly unlikely that the unobserved abilities of the station officers change at precisely the same time the derivative contracts are launched. The staggered nature of our shock makes it even less likely that our results are confounded by any such omitted factors. Furthermore, our maintained assumption is that the CME's selection of these derivative contracts is driven primarily by the demand for these hedging products at these cities, and not by *anticipated* improvement in the accuracy level of temperature measurement. Indeed, in unreported tests we find that hedging-demand considerations, such as electricity consumption and crop production in nearby areas, correlate well with the CME's decision to launch weather contracts on a specific city. Note that it poses no identification challenge for us if the CME chooses these stations based on their historical measurement accuracy. Station fixed effects separate out any such effect from our analysis. Our estimation comes from within-station changes in the accuracy level between the treatment and control stations.

As stated earlier, our use of the weather derivatives market is motivated by the empirical setting it provides us in teasing out the effect of market pressure on bureaucratic behavior. In other settings, such as the use of USDA data for the settlement of commodity futures contracts or the use of USGS data for assessing earthquake risk premiums, it is hard to pin down a specific shock that affects different cities, locations, or products in a staggered manner. It is also hard to come up with a very precise measure of inaccuracy in bureaucratic actions in these settings. The limitations of these other settings highlight the empirical advantages of our setting in terms of addressing the broader economic question.

B. Empirical Analysis

As a prelude to the main regression analysis, we provide the average number of error days for the treatment group for the three years before and after the shock and compare that to the corresponding averages for the control group. We plot these numbers in Figure 3.[21] The error rate drops slightly for the control group after the shock date, whereas there is a remarkable drop in the corresponding number for the treatment group. The average error rate drops from 12.86 to 11.29 days per year for the treatment group compared to a corresponding drop from 13.16 to 12.44 days per year for the control stations. In our regression model, we formally assess the statistical significance of the difference after removing the station and year fixed effects.

Results of the estimation exercise are provided in Table III. Columns (1) and (2) use the number of error days as the dependent variable, whereas columns (3) and (4) use the logarithm of the number of error days. Column (1) presents the results without year fixed effects. We obtain a coefficient of −2.36 on *Derivative*, indicating a decline of about 2.36 days in the annual error rate. The effect is statistically significant at the 1% level. In column (2), which includes the year fixed effects, we obtain a coefficient estimate of −1.63 that is also significant at the 1% level. Columns (3) and (4) obtain similar results and ensure that our estimates are not driven by outliers. These baseline results establish the effect of derivative introduction on measurement accuracy: NWS raw temperature readings become significantly more accurate once there is a direct financial market interest tied to these readings. In real terms, depending on the model specification, these estimates translate into a decline of about 13% to 20% in the error rate of the median station after the introduction of the derivative contracts. Our results are not confined to just a handful of stations—when we compute the difference in error rate on a station-by-station basis, we find that

[21] For this figure, we are unable to use the data for stations that received derivatives in the first wave (1999 to 2000). We do not have data for the past three years for these stations. All formal tests, presented in the rest of the paper, include these stations.

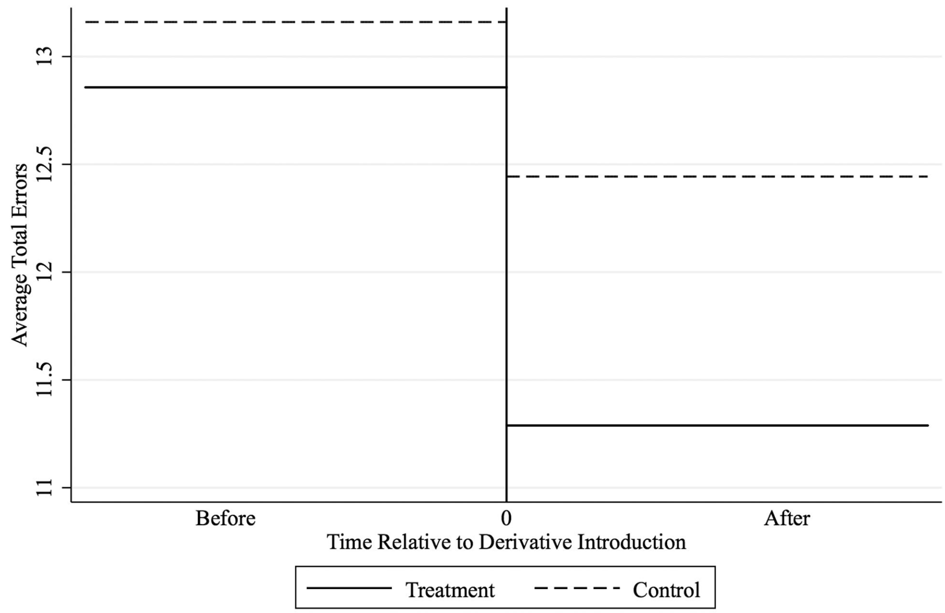

Figure 3. Average yearly errors pre- and postintroduction. This figure plots the average errors for the treatment and control groups before and after weather derivative introduction. The Before period is the three years before introduction and the After period is the year of introduction plus the three years afterward. The treatment group consists of the 14 weather stations that experienced a weather derivative introduction after 2002 and the control group consists of the 25 stations that never experienced a weather derivative introduction during our sample period.

18 of 20 treatment stations show improvement in measurement accuracy after the derivative introduction (see the Internet Appendix for details). Overall, these findings establish our base result: the introduction of financial markets improves the accuracy of the government agency's actions.

B.1. Differential Trend

We check for and rule out the presence of any preexisting declining trend in the error days of the shocked stations. We compute the change in error days from three years before to the year before derivative introduction and plot them in the Internet Appendix. Before the shock, both treatment and control stations show an increase, not a decrease, in the error rate. The shocked stations experienced an average change of +1% in the error rate during the preintroduction period as compared to the control stations. The difference in the rate of change for the two groups during the preshock period is statistically indistinguishable from zero (p-value = 0.95). After the introduction of the derivative, however, the treatment stations experience a considerable decline over the next three years. By the third postintroduction year, there is a decline of about 23% in the error rate of the shocked stations as compared

Table III

The Effect of CME Derivative Introduction on Weather Station Errors

This table presents results for our main regressions of weather station errors on CME derivative introduction. Observations are at the station-year level. The dependent variable in columns (1), (2), and (5) is the total number of weather station errors each year. The dependent variable in columns (3), (4), and (6) is the logarithm of the total number of weather station errors each year. *Derivative* is a dummy variable equal to one in the year of CME derivative introduction on the station and all years afterward. *Treat.* × *Trend* is a time trend for only treatment stations, zero otherwise. *Control* × *Trend* is a time trend for only control stations, zero otherwise. All regressions include station fixed effects. Columns (2), (4), (5), and (6) include year fixed effects. Standard errors clustered by weather station are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

	(1) Total Errors	(2) Total Errors	(3) Log (Total Errors)	(4) Log (Total Errors)	(5) Total Errors	(6) Log (Total Errors)
Derivative	−2.3577***	−1.6309***	−0.1828***	−0.1477***	−1.6953**	−0.1504**
	(0.46)	(0.58)	(0.04)	(0.05)	(0.69)	(0.06)
Treat. × Trend					−0.0475	−0.0020
					(0.10)	(0.01)
Control × Trend					−0.0625	−0.0026
					(0.10)	(0.01)
Observations	634	634	634	634	634	634
R^2	0.392	0.469	0.415	0.476	0.469	0.476
Year Fixed Effects	No	Yes	No	Yes	Yes	Yes
Station Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

to the corresponding decline for the control stations. These results show that there is no declining trend in measurement error of shocked stations before the launch, but a remarkable decline afterward. The improvement, therefore, is likely caused by the introduction of financial markets.

To formally rule out the impact of a differential trend on our results, we estimate our key empirical model with the inclusion of separate time trends for the treatment and control stations, *Treat.* × *Trend* and *Control* × *Trend*, respectively. The coefficient on *Derivative* in these regressions measures the effect of derivative introduction on the error rate after isolating the secular time trends across these two groups. The results are presented in columns (5) and (6) of Table III. The coefficients on both interaction terms are negative, indicating that the error rate has decreased over time across all stations. The estimated coefficients on *Treat.* × *Trend* and *Control* × *Trend* are statistically and economically indistinguishable from each other. More important, the coefficient on our variable of interest remains negative and significant. In column (5), the estimated coefficient on *Derivative* is −1.695, which is statistically significant at the 5% level. The corresponding coefficient in the base case regression presented in column (2) is −1.631. Similarly, the coefficient in the logarithmic error rate specification is comparable to the base case regression. Overall, these results show that our key finding is not driven by a differential trend across the treatment and control stations.

Table IV

The Effect of CME Derivative Introduction on Weather Station Errors: Derivative Stations Only

This table presents results for regressions of weather station errors on CME derivative introduction using only the sample of stations that eventually received a derivative contract during the sample period. Observations are at the station-year level. The dependent variable in columns (1) and (2) is the total number of weather station errors each year. The dependent variable in columns (3) and (4) is the logarithm of the total number of weather station errors each year. *Derivative* is a dummy variable equal to one in the year of CME derivative introduction on the station and all years afterward. All regressions include station fixed effects. Columns (2) and (4) include year fixed effects. Standard errors clustered by weather station are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

	(1) Total Errors	(2) Total Errors	(3) Log (Total Errors)	(4) Log (Total Errors)
Derivative	−2.3577***	−1.5580**	−0.1828***	−0.1409**
	(0.47)	(0.66)	(0.04)	(0.05)
Observations	311	311	311	311
R^2	0.322	0.450	0.340	0.460
Year Fixed Effects	No	Yes	No	Yes
Station Fixed Effects	Yes	Yes	Yes	Yes

B.2. Evidence from Treated Stations Alone

In our main test, we use a set of nonderivative stations as part of the control group. As we discussed earlier, the set of control stations is similar to the treatment stations along several important dimensions. However, there may be concerns that the nonderivative stations are not directly comparable to the derivative stations on some unobserved dimension. To rule out this possibility, we exploit the staggered nature of the introduction of contracts to estimate our main effect using the sample of 24 shocked stations only. Thus, stations that have not yet received a contract serve as counterfactuals for the stations that already have received contracts in any given year. Results are provided in Table IV. These results are almost identical to those reported using all 49 stations in our base case analysis. Stations show a decrease of 1.6 to 2.4 error days after the derivative launch. This test alleviates concerns that our results are driven by a specific set of control stations.

C. Economic Interests and Channels of Improvement

We argue that measurement accuracy improves due to increased economic interests and the resulting scrutiny of these measures by market participants. As economic interests increase, the reputational costs of measurement error are likely to increase as well. Our next set of tests exploits the cross-sectional variation in the level of economic interests across cities to provide evidence on this channel.

C.1. Cohort Analysis

We first relate economic interests to measurement improvement by exploiting variation in the year of contract introduction. Contracts introduced in earlier cohorts are likely to have higher economic interests as compared to later cohorts. This is based on the underlying assumption that the CME's incentive to introduce a weather contract in a city is driven primarily by the demand for weather hedging products in that city.

We separately estimate the effect of derivative introduction on measurement accuracy for each cohort as a first test of this hypothesis. We take all the shocked stations for a given cohort and include data from 1999 (i.e., the beginning year of the sample) to three years after the introduction year.[22] All the nonderivative stations during these years are included in the control group. Results are provided in Table V. We obtain a negative coefficient on *Derivative* for all four cohorts, with significant coefficients for all but one. Consistent with our hypothesis, the strongest effect comes from the earliest cohort (2000), whereas the weakest result comes from the last cohort (2008). For the 2000 cohort, we find a decline of almost six days per year in the error rate. The corresponding improvements are -2.0, -2.7, and -1.0 for the 2003, 2005, and 2008 cohorts, respectively.

As an additional empirical test of the cohort effect, we estimate the following model:

$$y_{st} = \alpha_s + \beta_{early} \cdot derivative_{st}^{early} + \beta_{late} \cdot derivative_{st}^{late} + year_t + \epsilon_{st}. \qquad (2)$$

In this model, $Derivative_{st}^{early}$ equals one for years after the derivative launch for all cities in the 1999/2000 cohorts, and zero otherwise, and $Derivative_{st}^{late}$ equals one for years after the derivative launch for all cities in the later cohorts, and zero otherwise. We estimate this model with data from the entire sample period. Results are provided in column (5) of Table V. The coefficient estimate of -5.1 on $Derivative_{st}^{early}$ is considerably larger than that of -1.2 on $Derivative_{st}^{late}$. The difference is statistically significant at the 11% level. This evidence is consistent with the idea that higher economic interests lead to higher visibility and better monitoring.

C.2. End-User Interest

The energy sector is the most important end-user of weather derivative products. Cities with high energy demand are therefore likely to have relatively higher interest in weather derivative products. We exploit this heterogeneity across weather derivative cities in our next test.

We take a city's population as a proxy for its energy demand. If a city falls among the top 25 population cities in the United States, then we classify it as a

[22] We limit the sample to three years postderivative introduction to estimate our main effect in the immediate aftermath of the launch. As an example, for the 2003 cohort, we include data from 1999 to 2006, and only the derivative stations that launched derivative contracts in 2003 (Kansas City, Houston, Boston, Minneapolis, and Sacramento) are included in the treatment group.

Table V

The Effect of CME Derivative Introduction on Weather Station Errors by Introduction Cohort and Hedging Demand

This table presents results for regressions of weather station errors on CME derivative introduction by cohort and by hedging demand. For each regression, we include only the derivative stations in the cohort of interest and all control stations. We run each regression from 1999 to three years after the cohort's introduction. The dependent variable is the total number of weather station errors for each station-year observation. The regressions in columns (1), (2), (3), and (4) include weather stations with CME derivative introduction in 1999 to 2000, 2003, 2005, and 2008, respectively. The regressions in columns (5) and (6) include all observations. *Early Cohorts* (*Late Cohorts*) is an indicator variable equal to one if a station is in the 1999 to 2000 (2003, 2005, or 2008) wave of introduction and has a derivative in that year. *High Demand* (*Low Demand*) is an indicator equal to one if a derivative is traded on the station in that year and the station is a high demand (low demand) location. High demand locations are locations with a population rank in the top 25 or in the first two cohorts (1999 and 2000). All regressions include station and year fixed effects. Standard errors clustered by weather station are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

	(1) Total Errors	(2) Total Errors	(3) Total Errors	(4) Total Errors	(5) Total Errors	(6) Total Errors
2000 shock	−5.7627**					
	(2.73)					
2003 shock		−1.9769**				
		(0.93)				
2005 shock			−2.7109**			
			(1.03)			
2008 shock				−0.9761		
				(0.81)		
Early Cohorts					−5.0916**	
					(2.33)	
Late Cohorts					−1.1895**	
					(0.50)	
High Demand						−2.2216***
						(0.72)
Low Demand						−0.6610
						(0.56)
Observations	201	263	287	400	634	634
R^2	0.604	0.524	0.511	0.476	0.473	0.471
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Station Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

high-energy demand location. Furthermore, consistent with the analysis of the previous section, we also classify a city as having high-energy demand if it gets a derivative contract in the first two waves (1999 and 2000). We create an indicator variable *High Demand* that equals one if the weather derivative station falls among the top 25 population cities or belongs to the early cohort. We similarly create an indicator variable *Low Demand* that equals $1 - $ *High Demand*. There are six cities in the *Low Demand* category. We expect these cities to have relatively lower economic interests and market scrutiny as compared to the remaining treatment cities in the sample.

We estimate our main regression model after including *High Demand* and *Low Demand* separately in the model. Results are provided in column (6) of Table V. We find that the improvements are concentrated in the subset of high demand cities. The coefficient estimate on *High Demand* is almost four times larger than that on *Low Demand*, at −2.22 as compared to −0.66. The difference is significant at the 6% level. Overall, these results are consistent with the notion that measurement outcomes improve with economic interests and market scrutiny.

C.3. Effort versus Investment

We now focus on the sources of temperature measurement improvement. There are two possible, not mutually exclusive, channels of improvement. First, the NWS might deploy more resources to the treatment stations, for example, by installing better thermometers and sensors, precisely when derivatives start trading. We call this the *investment* channel. Second, NWS officers might put forth more effort in measuring temperature after derivative introduction. We call this the *effort* channel. Effort-related improvements would come through better monitoring, calibration, and maintenance of the weather stations to minimize erroneous reports. We provide two empirical tests in this section that lend support to the effort channel. In the first test, we exploit interesting seasonal variation in weather derivative trading activity. In the second test, we directly investigate the extent of investment made at different locations by the NWS.

Seasonal Test: We empirically attempt to separate the effort and technology channels by exploiting variation in trading activities in the weather derivatives market across calendar months. This test also allows us to strengthen our claim that economic interests drive more accurate measurement. The end-users of the weather derivatives market are typically sectors such as utilities, farming, transportation, retail, and food products. A majority of their hedging demands arise in the months with more extreme temperatures. Not surprisingly, we find that an overwhelming majority of these contracts are traded in peak summer and peak winter months. This leaves the months of April and October as the least traded months on the exchange. We estimate the basic regression model separately for these two months and the rest of the year. The key idea is to assess the improvement in measurement effort, keeping the underlying measurement technology the same.

We aggregate all the error days in April and October for the first analysis and, similarly, all the error days in the remaining months for the second analysis. Results are provided in columns (1) and (2) of Table VI, respectively. In column (1), we find no improvement in October and April, whereas in column (2) there is significant improvement in the active trading months. In an additional test, we separate the sample into two groups by placing the three months around April and the three months around October in one group, and the rest in another. Thus, we have exactly six months in each group. As shown in columns (3) and (4) of the table, negative coefficients obtain for both groups, but the

Table VI

Monthly Market Activity and the Effect of CME Derivative Introduction on Weather Station Errors

This table presents results from regressions of weather station errors on CME derivative introduction where the sample is separated into active and inactive months. The dependent variable is the total number of weather station errors in the months of interest for each station and year. The regression in column (1) *excludes* all months except April and October, the least active months based on open interest. The regression in column (2) *includes* all months except April and October. The regression in column (3) *includes* the months of April and October and the months on each side of April and October. The regression in column (4) *excludes* the months of April and October and the months on each side of April and October. *Derivative* is an indicator equal to one if a derivative is traded on the station in that year and all years afterward. All regressions include year and station fixed effects. Standard errors clustered by weather station are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

	(1) Total Errors	(2) Total Errors	(3) Total Errors	(4) Total Errors
Derivative	0.0546	−1.7142***	−0.4829	−1.1637***
	(0.25)	(0.58)	(0.42)	(0.42)
Observations	634	634	634	634
R^2	0.211	0.445	0.374	0.355
Year Fixed Effects	Yes	Yes	Yes	Yes
Station Fixed Effects	Yes	Yes	Yes	Yes
Active/Inactive Months	Inactive	Active	Inactive	Active

coefficient is significant only for the peak months. Furthermore, the coefficient for the peak months subsample is more than double that for the off-peak months. Thus, at a given station, the improvement comes from months with active trading interest. These are the months when the pressure and scrutiny from the outside market are likely to be highest—in these months, the frequency of follow-ups with the NWS stations and analysis of the weather data by trading professionals is expected to be higher than the remaining months. Our results thus support the view that financial markets induce higher effort by the employees in measuring temperature accurately.

Upgrades and Improvements: One of the most important recent changes in the NWS's temperature measurement technology was the installation of ASOS. Although the NWS consistently upgrades these stations, the installation of ASOS was the most important event on the technology side of temperature measurement. We collect data on the ASOS installation dates for 48 of the 49 weather stations in our sample.[23] For each of these stations, ASOS was commissioned before September 1999. This rules out the possibility of any major change in measurement technology just after the introduction of the derivative contract. In addition, we correlate the year of introduction of ASOS with an indicator variable that equals one for the treatment group and zero for the control group. The correlation coefficient is almost zero. Thus, we do not find any evidence that derivative stations in our sample receive major

[23] The data for the Los Angeles station are not available.

technological upgrades earlier than the control stations. However, this evidence does not preclude incremental investments at the treatment stations after the contracts were introduced. We analyze the patterns of incremental investments across treatment and control stations over time to shed light on this channel.

The incremental improvements undertaken by NWS during our sample period can be broadly categorized into three groups: (i) improvement of software, (ii) upgrade of processor board, and (iii) installation of upgraded hardware in the ASOS. We obtain data on these upgrades from NOAA and compare them across treatment and control stations. In particular, we assess whether the treatment stations obtained better upgrades or earlier incremental investments in response to the weather derivative launch as compared to the control stations.

We first describe these improvements and the data source in detail. We then conduct an analysis of the difference across the two groups.

Software Upgrade: ASOS use various pieces of software for properly recording and transmitting weather data. Over time, NOAA has undertaken a series of improvements in the software used by these systems. Different versions of software are clearly identified with numbers such as 2.6 or 2.7. We are able to obtain data on the software versions used at various weather locations in 2003. Ideally, we would use panel data on the upgrades over time and across stations, but we are not able to get such data from NOAA. Our cross-sectional data, however, allow us to test for the similarity of software versions at a crucial midpoint in our sample period. If we find no difference in software versions across the two groups in 2003, it is unlikely that there was a systematic difference in these two groups' software technology.

Processor Board Improvement: Subsequent to the ASOS installation, NOAA made a significant investment in replacing the processor board of the unit. The upgrade consisted of a new processor board with expanded memory and processing capability to handle sensors with higher speed, greater reliability, and communication enhancements. Weather stations were divided into different implementation groups based on a host of criteria such as their overall importance and other technical and budget-related factors. The implementation of the processor board was carried out based on these groupings. We obtain data on the assignment of stations into these implementation groups to check whether the treatment stations obtained this upgrade earlier than the control stations.

Other Hardware Installation: Finally, there have been other upgrades and improvements in some sensors and the monitoring gauge of the ASOS system. These improvements include the installation of Dew Point Temperature Sensors (DTS), Ice Free Wind Sensor Installation (IFW), and All-Weather Precipitation Accumulation Gauge (AWPAG). The exact installation dates of these sensors are available from NOAA. We use these improvement dates to test for differences in the timing of these changes across the two groups.

We conduct three empirical tests to assess the differences in technological investments across treatment and control stations. Results are provided in Table VII. Panel A provides the cross-sectional difference in software versions

Table VII

Weather Station Technological Improvements

This table presents information and tests on weather station technological improvements during the sample period. Panel A presents the 2003 cross-sectional distribution of automated surface observing system software (ACU S/W) and the automated surface observing system planned processor improvement group (Proc. Group). Below each distribution we provide the p-value of a Pearson's χ^2 test. Panel B presents the mean year of technological improvement for the derivative and treatment stations. The three improvements are an ice free wind sensor (IFW), dew point sensor replacement (DTS), and all-weather precipitation accumulation gauge (AWPAG). Column (4) gives the difference in means. Panel C presents results from regressions of technological improvement on derivative introduction year. Observations are at the station-year level. The dependent variables are dummy variables equal to one if the weather station received the technological improvement during the year. In column (1), the technological improvement is *IFW*. In column (2), the technological improvement is *DTS*. In column (3), the technological improvement is *AWPAG*. *Intro Year* is a dummy variable equal to one if the weather station received a derivative during the year. All regressions include year fixed effects. Standard errors clustered by weather station are in parentheses. ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$.

Panel A: 2003 Technology Cross-Section

ACU S/W	Control	Derivative
2.6	16	16
2.63	4	4
2.7B-6	5	4
χ^2 p-value = 0.956		

Proc. Group	Control	Derivative
6	3	3
7	18	16
8	0	1
9	4	4
χ^2 p-value = 0.778		

Panel B: Technology Improvement Years

Enhancement	Control	Derivative	Difference
IFW	2007.320	2007.292	0.028
DTS	2004.560	2004.292	0.268
AWPAG	2004.696	2004.292	0.404*

Panel C: Technological Improvement Regressions

	(1) IFW	(2) DTS	(3) AWPAG
Intro Year	0.0104	0.0199	0.0019
	(0.04)	(0.06)	(0.05)
Observations	637	637	637
R^2	0.241	0.263	0.366
Year Fixed Effects	Yes	Yes	Yes
Station Fixed Effects	No	No	No

and the processor board implementation groups of the treatment and control stations. The distribution of the version of software used by the two groups is almost identical. A χ^2 test for the equality of distribution of software versions across the two groups cannot be rejected (p-value=0.96). Similarly, the distribution of the processor group is almost identical across the two groups (p-value=0.78).

Panel B provides the average year of the implementation of DTS, IFW, and AWPAG across the two groups. In all three cases, the average year of implementation is the same across the two groups—2007 for IFW, 2004 for DTS, and 2004 for AWPAG. Note that on average these changes happened after the introduction of derivatives in the earlier cohorts, that is, cohorts with the maximum improvement in measurement error. Thus, it is unlikely that these changes have any meaningful impact on our main findings. However, even if they did have some impact, it would not be likely to affect the two groups differentially since they received these improvements in the same year. None of the differences are significant at the 5% level.

In Panel C, we examine if weather derivative stations received these improvements around the time of derivative introduction. This test allows us to investigate if derivative launch was accompanied by technological changes. The dependent variable is one for the year of technological improvement and zero in the other years. For example, Atlanta obtained an IFW in 2007. The dependent variable for Atlanta is one for 2007 and zero otherwise. The explanatory variable is an indicator variable equal to one in the year of derivative launch and zero otherwise. Columns (1) to (3) provide the regression results for IFW, DTS, and AWPAG, respectively. We include year fixed effects to separate the effect of derivative launch from national upgrades undertaken by NOAA. We find no evidence that technological improvement happened during the year of derivative introduction. In the Internet Appendix, we repeat this exercise with the year before the derivative launch and the year after the derivative launch as explanatory variables to examine whether these stations received better technology just before or just after the launch. Again, there is no material difference between the two groups.

In summary, we find no meaningful difference in the level of technology at treatment and control stations. In light of limited data and the complexity of the task, we acknowledge our limitation in precisely establishing a channel of improvement. However, our results are broadly consistent with the effort-based channel.

These tests also allow us to comment on an alternative interpretation of our results. If the improvement comes from greater investment in technical resources, then our findings are less likely to be based on market discipline. One may argue that the government agency was optimally investing few resources in temperature measurement before the launch of the contract, and that, after the contract launch, with more at stake on the temperature measurement, the agency optimally deploys more funds to these stations. In effect, the agency may be moving from one efficient outcome to another. In the absence of detailed data on all potential costs incurred in achieving more accurate temperature measurement, we are limited in our ability to completely rule out this alternative interpretation. However, our results show that NWS did not invest more resources in the technical capabilities of the treated stations. Despite having similar technical resources as the control stations, the treated stations obtain more accurate measurement. Therefore, we are able to rule out arguably the most important driver of the alternative interpretation, namely, higher investment by the NWS at these locations.

D. Economic Magnitude

We show that the treatment stations improve their measurement accuracy by 13% to 20% after the launch of derivative contracts. How important are these improvements in economic terms? We comment on the economic implications of our results for three main stakeholders in our setting: financial market participants, NWS, and society as a whole.

D.1. Economic Implications for Financial Market Participants

Implications for Speculators: We compute the hypothetical difference in the payoffs of these contracts if they were to settle on the raw versus corrected temperature numbers. The average weather station has about one error-day per month, with an average magnitude of 3°F. Since each degree day is worth \$20 as per the contract, a 3°F error translates into a payoff difference of \$60 for a monthly contract. Thus, the unconditional error rate represents about 1.58% of the notional contract value (\$60/\$3,800).[24] An improvement of 13% to 20% in the error rate translates into a difference of 21 to 32 basis points of the monthly contract's notional value (13% to 20% of 158 basis points). As shown earlier, our results are concentrated within the early cohorts, which have larger economic interests in these markets. The difference between the raw and corrected temperature number translates into a difference of 65 basis points for stations in the early cohorts. Thus, the volume weighted economic impact is likely to be even higher than the average effect of 21 to 32 basis points. At an aggregate level, we estimate the total dollar magnitude of this effect at approximately \$47 million for the early cohorts alone. We provide details on the computation of these numbers in the Internet Appendix.

Overall, the economic magnitude of the effect is modest in the context of broad financial markets. However, a difference of 21 to 32 basis points is likely to be relevant for financial institutions that are highly levered. Since financial institutions take positions in this market based on margin requirements of 4% to 17% (see the Internet Appendix), the payoff difference magnifies to 0.78% to 4.4% in terms of deployed capital. Finally, given the small number of investors

[24] We estimate the average notional value of monthly contracts at about \$3,800 during our sample period. The notional value of a contract is computed as two times the standard deviation of the historical degree day index as per the guidelines of WRMA. For the monthly contracts in our sample, this is approximately \$3,804.

in the weather derivatives market, the aggregate difference of approximately \$47 million is likely to be meaningful on a per investor basis.

Implication for Hedgers: Corporations are less likely to engage in hedging activities if the hedging instruments have higher basis risk (e.g., see Haushalter (2000)). Thus, the improvement in measurement accuracy should benefit hedgers such as utilities and energy companies by decreasing the basis risk of weather derivative instruments. To provide some evidence in support of this claim, we compare the open interest in the weather derivatives market across stations with varying levels of error rate. Compared to stations that fall in the top nine (least accurate measurement) in terms of error rate, stations in the bottom nine (most accurate) have 47% higher open interest. The corresponding difference is 54% when we compare the top six with the bottom six stations. These results suggest that concerns about error rate impede the development of the market. A decrease in basis risk due to better temperature measurement, therefore, is likely to have a positive effect on hedging efficiency. This can attract more participants to the market, which in turn can increase the liquidity of the market. Hence, there is the possibility of a positive feedback effect from the documented measurement improvements.[25]

D.2. Reputational Cost to NWS

We argue that the increased visibility and scrutiny by market observers creates significant reputational concerns for the government agency. This scrutiny, in turn, increases the reputational cost of a mistake. It is hard to quantify reputational costs in dollar terms. Instead, we provide some qualitative evidence to support the claim that the NWS paid significant attention to this concern. First, we show that NWS issued directives to its field offices highlighting the increased scrutiny by media and private industry in recent years. Second, we document that, in the early stages of the development of this market, NOAA staff and the representatives of the weather derivatives market held meetings to discuss the viability of the market and NWS's crucial role in measuring temperature data.[26] We discuss this evidence in detail in the Internet Appendix.

D.3. Welfare Implications

Since we do not observe all the costs and benefits of the improvement in measurement accuracy, we are unable to make any welfare claims based on our

[25] Anecdotal evidence and practitioners' reports lend further support to the importance of basis risk in the weather derivatives market. For example, Tindall (2006) argues that rainfall derivatives are less common than temperature derivatives because of the higher basis risk of rainfall contracts. Buckley et al. (2002) discuss data quality as a reason behind the relative lag in development of the European weather derivatives market in comparison to the U.S. market. Industry practitioners also highlight the role of basis risk as one of the key reasons for a lower participation rate by investors and hedgers in hedging markets (e.g., the report on effectiveness of hedging by the American Academy of Actuaries (1999)).

[26] Weather scientists have also taken note of the increased attention paid to climate observations by the private sector in recent years (e.g., see Changnon and Changnon (2010)).

Table VIII

The Effect of CME Derivative Introduction on Precipitation Errors

This table presents results for regressions of precipitation recording errors on CME derivative introduction. Observations are at the station-year level. The dependent variable is the total number of errors during the year. *Derivative* is a dummy variable equal to one in the year of CME derivative introduction on the station and all years afterward. The regression includes station and year fixed effects. Standard errors clustered by weather station are in parentheses. ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$.

	Precipitation Errors
Derivative	−0.3849
	(0.50)
Observations	441
R^2	0.157
Year Fixed Effects	Yes
Station Fixed Effects	Yes

study. Pérez-González and Yun (2013) analyze the beneficial effects of weather derivative contracts for energy utilities. They show that hedgers can lower their cost of capital and bankruptcy risk by using these contracts. As discussed earlier, we find some evidence that the level of activity in the weather derivatives market is positively correlated with the accuracy of temperature measurement. Thus, if the improvement in measurement accuracy comes at little or no cost, it can potentially have some welfare implications.

E. Robustness Tests

E.1. Falsification Test Using Nontemperature Measures

In addition to temperature, weather stations collect several other pieces of information such as precipitation rates. The temperature-based contracts do not depend on the accuracy of these measures. Hence, as a placebo test, we estimate the effect of temperature contract introduction at a station on the accuracy of the precipitation rate at the same station. This test allows us to rule out whether some confounding variables, and not the financial market scrutiny of the temperature data, drive our main results.

We obtain precipitation data for all 49 weather stations that enter our main sample. Unlike temperature data, we are unable to obtain a measure of raw and corrected precipitation rate. We do, however, observe if a given station missed the precipitation recording at a given location on a given day. Thus, we have a measure of precipitation data accuracy based on extreme observations—the nonavailability of the measurement itself. We reestimate our base model using the number of missing precipitation observations in a year as the dependent variable. As reported in Table VIII, the estimated coefficient on *Derivative* is insignificant, suggesting no change in the accuracy of this measure around the contract launch date.

E.2. NCDC Cleaned Values

All of our results so far are based on the difference between a third-party (MDA) certified measure of clean data and the NWS raw data for a station's temperature. We also obtain the corrected or restated data produced by an affiliated government agency of the NWS, namely, the NCDC. The NCDC is responsible for producing the government's final data after removing measurement errors by the station.

We reestimate our main regression models based on the NCDC data and report the results in the Internet Appendix.[27] The results are almost identical to those reported using MDA values. These results provide confidence in our measure of temperature accuracy since data from both parties—MDA, a third-party private company, and NCDC, an affiliated government agency—produce similar results.

E.3. Controlling for Changes in Weather Conditions

It is unlikely that weather conditions become more conducive to better measurement outcomes after the launch of derivative contracts in a city. If that were not the case, then an improvement in measurement accuracy could simply be an artifact of changes in weather conditions. To rule out this possibility, we reestimate our main empirical model after including the volatility of annual temperature and the level of temperature as additional explanatory variables in the model. Results are reported in the Internet Appendix. We find that the coefficient on *Derivative* remains negative and significant, and similar to the base case estimate in economic terms.

E.4. Quarterly Analysis

Our main tests are based on annual observations. In the seasonality-based test, we also exploit the seasonal variation across active and passive months of trading. As a robustness test, we estimate our model based on quarterly observations. In this test, we consider each station-quarter as a unit of observation and reestimate our base model. We include station-calendar quarter and year fixed effects in the model to soak away variation specific to these levels. We find that the effect of contract introduction on temperature accuracy starts in the quarter of introduction and continues to increase over the next five to six quarters. After this time, the effect flattens out (see the Internet Appendix).

[27] We begin with all 49 stations in our sample. However, we do not have high-quality NCDC data for 1999. In addition, the agency did not produce corrected values for two control stations (San Jose and Riverside) during our sample period. Hence, we lose 1999 and two control stations from the sample. Also, the NCDC did not produce corrected values for eight stations in December 2001, so we lose one observation month for this part of the study.

V. Conclusion

Government agencies and financial markets interact with each other in a number of settings. A clear understanding of the effect of financial markets on the incentives of bureaucracies thus has important implications for the economy. We choose the specific setting of the weather derivatives market to analyze this effect. Our setting allows us to cleanly identify the effect of market pressure on the relevant government agency's actions.

We show that the launch of a weather derivatives market on a city's temperature results in more accurate temperature measurement by the dedicated weather station for that city. After the launch of these contracts, the NWS-reported numbers become reference points for a large amount of contracts in the private market. Thus, there is increased interest in and scrutiny of these numbers by third parties, which in turn creates more pressure on the NWS to produce better measures. The increased pressure can come in the form of potential reputational loss or the possibility of future disputes among the contracting parties. We show that our results are economically meaningful, and provide enough incentives to speculators, hedgers, and the exchange to subject NWS to higher scrutiny. Our results highlight an important, but relatively unexplored, role of financial markets: they can alter the actions of bureaucracies even in the absence of explicit incentives and monitoring mechanisms.

Initial submission: May 22, 2013; Final version received: September 23, 2015
Editor: Michael Roberts

REFERENCES

Admati, Anat R., and Paul Pfleiderer, 2009, The "Wall Street Walk" and shareholder activism: Exit as a form of voice, *Review of Financial Studies* 22, 2645–2685.

Allayannis, George, and James P. Weston, 2001, The use of foreign currency derivatives and firm market value, *Review of Financial Studies* 14, 243–276.

Allen, Franklin, and Douglas Gale, 1994, *Financial Innovation and Risk Sharing* (MIT Press, Cambridge, MA).

American Academy of Actuaries, 1999, Evaluating the effectiveness of index-based insurance derivatives in hedging property/casualty insurance transactions, Index Securitization Taskforce Report. Available at http://www.casact.org/research/istf/istf.pdf

Becht, Marco, Patrick Bolton, and Ailsa Röell, 2003, Corporate governance and control, in George M. Constantinides, Milton Harris, and René M. Stulz, eds.: *Handbook of the Economics of Finance* (Elsevier, North Holland).

Berrospide, Jose M., Amiyatosh Purnanandam, and Uday Rajan, 2010, Corporate hedging, investment and value, Working paper, University of Michigan.

Black, Bernard S., 1998, Shareholder activism and corporate governance in the U.S., in P. Newman, ed.: *The New Palgrave Dictionary of Economics and the Law* (Palgrave Macmillan London, UK).

Boardman, Anthony E., and Aidan R. Vining, 1989, Ownership and performance in competitive environments: A comparison of the performance of private, mixed and state-owned enterprises, *Journal of Law and Economics* 32, 1–33.

Bolton, Patrick, and Ernst-Ludwig Von Thadden, 1998, Blocks, liquidity, and corporate control, *Journal of Finance* 53, 1–25.

Buckley, Nick, Alex Hamilton, Justyn Harding, Nicola Roche, Nick Ross, Emma Sands, Richard Skelding, Nick Watford, and Heidi Whitlow, 2002, European weather derivatives, Paper presented at General Insurance Convention 2002.

Burkart, Mike, Denis Gromb, and Fausto Panunzi, 1997, Large shareholders, monitoring, and the value of the firm, *Quarterly Journal of Economics* 112, 693–728.

Carter, David A., Daniel A. Rogers, and Betty J. Simkins, 2006, Does hedging affect firm value? Evidence from the U.S. airline industry, *Financial Management* 35, 53–86.

Changnon, David, and Stanley A. Changnon, 2010, Major growth in some business-related uses of climate information, *Journal of Applied Meteorology and Climatology* 49, 325–331.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, 1999, The economics of career concerns, Part II: Application to missions and accountability of government agencies, *Review of Economic Studies* 66, 199–217.

Dutton, John A., 2002, Opportunities and priorities in a new era for weather and climate services, *Bulletin of the American Meteorological Society* 83, 1303–1311.

Edmans, Alex, 2009, Blockholder trading, market efficiency, and managerial myopia, *Journal of Finance* 64, 2481–2513.

Gillan, Stuart, and Laura Starks, 1998, A survey of shareholder activism: Motivation and empirical evidence, *Contemporary Finance Digest* 2, 10–34.

Gopalan, Radhakrishnan, 2009, Institutional stock sales and takeovers: The disciplinary role of voting with your feet, Working paper, Washington University, St. Louis.

Haushalter, G. David, 2000, Financing policy, basis risk, and corporate hedging: Evidence from oil and gas producers, *Journal of Finance* 55, 107–152.

Hayek, Friedrich A., 1945, The use of knowledge in society, *American Economic Review* 35, 519–530.

Heckman, James, Carolyn Heinrich, and Jeffrey Smith, 1997, Assessing the performance of performance standards in public bureaucracies, *American Economic Review*, 87, 389–395.

Hermalin, Benjamin E., and Michael S. Weisbach, 2003, Boards of directors as an endogenously determined institution: A survey of the economic literature, *Economic Policy Review* 9, 7–26.

Holmström, Bengt, and Jean Tirole, 1993, Market liquidity and performance monitoring, *Journal of Political Economy* 101, 678–709.

Kahn, Charles, and Andrew Winton, 1998, Ownership structure, speculation, and shareholder intervention, *Journal of Finance* 53, 99–129.

Karpoff, Jonathan M., 1998, Impact of shareholder activism on target companies: A survey of empirical findings, Working paper, University of Washington.

Karpoff, Jonathan M., 2001, Public versus private initiative in Arctic exploration: The effects of incentives and organizational structure, *Journal of Political Economy* 109, 38–78.

Pérez-González, Francisco, and Hayong Yun, 2013, Risk management and firm value: Evidence from weather derivatives, *Journal of Finance* 68, 2143–2176.

Prendergast, Canice, 2007, The motivation and bias of bureaucrats, *American Economic Review* 97, 180–196.

PricewaterhouseCoopers, 2006, Weather Risk Management Association Survey.

Purnanandam, Amiyatosh, 2007, Interest rate derivatives at commercial banks: An empirical investigation, *Journal of Monetary Economics* 54, 1769–1808.

Romano, Roberta, 2001, Less is more: Making institutional investor activism a valuable mechanism of corporate governance, *Yale Journal on Regulation* 18, 174–251.

Shleifer, Andrei, and Robert W. Vishny, 1986, Large shareholders and corporate control, *Journal of Political Economy* 94, 461–488.

Shleifer, Andrei, and Robert W. Vishny, 1997, A survey of corporate governance, *Journal of Finance* 52, 737–783.

Stulz, René M., 2004, Should we fear derivatives? *Journal of Economic Perspectives* 18, 173–192.

Tindall, Jon, 2006, Weather derivatives: Pricing and risk management applications, Institute of Actuaries of Australia for the Financial Services Forum, Australia. Available at http://actuaries.asn.au/Libarary/fsfob_paper_tindall_weather%20derivatives.pdf

Tufano, Peter, 2003, Financial innovation, in George M. Constantinides, Milton Harris, and René M. Stulz, eds.: *Handbook of the Economics of Finance* (Elsevier, North Holland, Amsterdam).

Wilson, James Q., 1989, *Bureaucracy: What Government Agencies Do and Why*? (Basic Books, Inc., New York, NY).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.