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

(Preliminary Draft)

Amiyatosh Purnanandam† Daniel Weagley‡

October 18, 2012

Abstract

The payoffs of weather derivative contracts depend on variables such as temperature and snowfall levels in the underlying city during the contract period. Chicago Mercantile Exchange (CME) has introduced several temperature related contracts on different U.S. cities in a staggered fashion over the past 13 years. We show that the introduction of these contracts on a city's temperature improves the accuracy of temperature measurement by the dedicated weather station in that city in a causal manner. We argue that the introduction of temperature-based financial markets generates additional scrutiny of the temperature data measured by the National Weather Services (NWS), which in turn produces better outcomes by the government agency. Our results have important implications for the role of financial innovation and markets in affecting real outcomes.

*We thank Taylor Begley, Sugato Bhattacharyya, Ing-Haw Cheng, Paolo Pasquariello, Maciej Szefler, Uday Rajan, and seminar participants at the University of Michigan for helpful comments. All remaining errors remain our responsibility.

†Ross School of Business, University of Michigan; amiyatos@umich.edu

‡PhD Candidate, Ross School of Business, University of Michigan; weagley@umich.edu

1 Introduction

Weather has a large impact on a variety of economic and social decisions. Accurate and timely measurement of weather, therefore, is a high priority task for several government agencies around the globe.[1] As expected, there has been a number of technological improvements over the years to meet the need for better measurement of weather variables such as temperature and rainfall. In this paper, we focus on a different mechanism, namely the role of financial markets in improving the accuracy of temperature measurement by the National Weather Services (NWS), the main government agency responsible for collection and dissemination of weather data to the private and public sectors.

While the need for managing weather-risk has been felt for centuries, it was only in 1999 that financial markets began to offer exchange traded weather derivatives contracts to protect against weather related financial risk. Since then, the Chicago Mercantile Exchange (CME) has introduced weather derivative contracts on a number of U.S. cities over time in a staggered manner. As per the survey results of the Weather Risk Management Association (WRMA), an industry body, more than 95% of these contracts in terms of notional value are temperature related (WRMA, 2006). The temperature contracts are settled based on daily temperature data measured by the underlying city's weather stations. These weather stations are operated by the National Weather Service (NWS) or affiliated government agencies. Thus the introduction of these contracts on the financial markets directly ties the NWS reported temperatures to the very large economic interests of traders and hedgers in this market. We argue that such a direct linkage between the NWS weather measure and large financial interests leads to a tremendous scrutiny of the NWS's actions by financial investors and related parties. The resulting visibility and market scrutiny works as a disciplining device for the NWS by motivating them to minimize measurement errors that can arise due to factors such as improper calibration, monitoring, and maintenance of the measurement equipment.

It has been well recognized that financial markets can work as a disciplining device for cor-

[1]The U.S. alone has 15 federal departments or agencies working on meteorological issues.

porate managers whose incentives are either directly or indirectly tied to the firm's stock performance (e.g., see Holmstrom and Tirole (1993), Shleifer and Vishny (1986), and Kahn and Winton (1998)). There is no such direct monetary incentive mechanism or monitoring pressure from the financial markets on the NWS officers in our setting. Our results, therefore, are some of the first pieces of evidence that financial markets can produce better real outcomes even in the absence of any explicit incentive and monitoring mechanism in place. Why should the NWS care about producing better outcomes in the absence of explicit incentive mechanisms? We argue that there are at least two important reasons behind this: (a) the possibility of reputation loss, and (b) the avoidance of future disputes and lawsuits.[2] As the NWS-reported temperature numbers become reference points for the settlement of a large volume of financial contracts, discrepancies in these numbers are more likely to create reputation damages for the agency. If a public agency experiences a loss in reputation they may be subject to political hearings and downsizing. Noted social scientist James Q. Wilson observes "The head of a business firm is judged and rewarded on the basis of the firm's earnings–the bottom line. The head of a public agency is judged and rewarded on the basis of the *appearance* of success, when success can mean reputation, influence, charm, the absence of criticism, personal ideology, or victory in public debate" (Wilson (1989), page 197).

Related to the reputation concerns, there is a higher probability of disputes arising out of improper recording of the temperature since an error can now cause immediate and direct financial loss to third parties. Even though the government agency may not be a party in resulting litigations, they may experience negative publicity or a loss of reputation due to the lawsuit. We provide a number of pieces of descriptive evidence, collected from a variety of sources such as NWS's directives and the industry interest group's documents, in support of these channels. Our evidence is also consistent with the well-known "Hawthorne effect" where experimental subjects increase their productivity when they are aware they are under observation.[3]

[2]There is a large literature, both theoretical and empirical, on the effect of reputation on financial contracting in private markets. For example, Diamond (1989) studies reputation formation in debt markets.

[3]See Levitt and List (2009) for the discussion of the effects as well as some recent analysis and counter arguments about the original Hawthorne data set.

As of December 31, 2011, there are 24 U.S. cities with temperature related derivative contracts trading on them. These contracts were issued in four different waves in 1999-2000, 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 non-shocked stations (the control group) during the same period. The staggered nature of the derivative launch allows us to separate the effect of any time trend in error rate or any general improvement in NWS's technology over time. Our empirical setting has another important advantage in terms of establishing causality. Unlike stocks, bonds, or foreign currencies, the variable underlying the weather derivatives contract is not a traded commodity by itself. Thus the introduction of the derivative contract is not going to effect the value of the underlying assets – a concern that is always present in studies that analyze the role of derivative contracts on the underlying assets. This property ensures that we do not suffer from any reverse causality concern from the derivatives market to the value of the underlying.

To measure weather station accuracy we compare two sets of temperatures. The initial record of temperature for each weather station is issued by the NWS in a standard format called METAR reports. These numbers are often called the *raw* or *preliminary* temperatures. These temperature readings sometimes contain errors due to reasons such as equipment malfunction, improper installation of the equipment, or improper calibration and maintenance of the station.[4] These data errors ultimately result in either incorrect temperature records or missing values. We obtain data on the *true* or *cleaned* values of temperature from two sources and take the discrepancy between the raw and cleaned values as the measure of measurement error. The first source of *cleaned* temperature values is a private company called MDA Information Systems Inc. that specializes in correcting the raw temperature data from the NWS. They use a number of techniques to clean up the raw data including recovering data from alternative sources, using their proprietary model to correct mistakes, cross-checking the NWS data against other nearby stations and by calling up the climate centers including NWS field offices to discuss possible

[4]For example, see NWS instruction number 10-1302 or NWS 10-1004 for steps undertaken by the weather stations to minimize errors in data gathering exercise.

errors. The second source of cleaned data is an affiliated government agency of the NWS itself, called the National Climatic Data Center (NCDC). At the time of its initial report, NWS makes it clear that the initial temperature numbers are preliminary and subject to change based on their data cleaning and verification exercise. After their verification exercise they report revised temperature numbers. Based on these two measures of cleaned values, we are able to compute the extent of measurement error in the initial data collection exercise by the NWS weather stations. Our results remain similar based on either measure of correct value since NCDC and MDA reported clean temperature values are almost identical.

Our estimation shows that after the introduction of weather derivatives, the shocked station's error rate comes down by about 10%. Said differently, these stations have lower incidence of inaccurate data after their recorded temperature numbers become reference points for billions of dollars of financial contracts in an open market. The results are significant both in a real and statistical sense. We further show that the effect is strongest among stations where economic interests are likely to be higher. We do so by first showing that the improvement in measurement accuracy mostly comes from stations with high open interest. Second, the effects are stronger for cities with relatively higher populations. Overall these results establish our main claim that the launch of weather derivatives results in better measurement outcomes by the NWS and the result is most likely linked with financial interests generated through derivative contracts.

It may be possible that the NWS improves its technology at the derivative launch stations precisely at the time of derivative introduction. In such a case, the effect that we document would come solely from the improvement in technology at these stations rather than through better effort by NWS employees. Both these channels, better technology and better effort, are consistent with our main results that financial markets affect real outcomes and work as a disciplining device. However, we are able to separate the two by exploiting a unique seasonality feature of this market. An overwhelming majority of these contracts are traded to protect against high heat and extreme cold conditions. The Heating Degree Day contracts (HDD) are used by the hedgers in the winter months to hedge against cold weather. The Cooling Degree Days

(CDD) contracts are used in summer months to protect against hot weather. There are two months of the year, called the "cross-over months" by many market participants, when there is very little activity in either contract's market. These are the months of April and October. If the NWS selectively introduces better equipment at these stations at the time of derivative launch then the improvement in measurement accuracy should be felt throughout the year. If, on the other hand, better effort is put forth by officials when economic interests are high, then we expect to see higher improvement in peak months and not much of a difference in April and October. Our results support the latter interpretation. All the accuracy improvements come from months excluding April and October, and there is no change in the measurement accuracy in these two months. As a robustness exercise, we widen the cross-over period to six months including months immediately preceding and following April and October. We show that our results mainly come from the peak activity months of June-August and December-February, i.e., from a period of high economic interest in this market. Our evidence supports the view that there has been an improvement in the accuracy rate due to better effort in maintenance and monitoring of these stations, conditional on the technology available at the station.

Our results have important implications for the role of markets in improving efficiency. While there has been a volume of research on the role of markets in improving the allocative efficiency of the economy, little is known empirically about the role of markets in improving government agencies' actions. Related, our evidence shows that while governments often regulate markets for better behavior by market participants, markets can regulate government through the channels of increased scrutiny and visibility. Second, our study directly relates to the role of financial innovations and derivatives on real outcomes.[5] The extant literature has made good progress in analyzing the role of derivatives markets on the corporations that use them (see Perez-Gonzalez and Yun, 2012). Our study provides a unique perspective on the role of financial innovations on real outcomes by documenting some benefits of these markets that accrue due to the actions of third parties, i.e., parties not directly participating in the markets. To the extent a more

[5] See Tufano, 2003 for a survey on financial innovation including the role of innovation on society.

accurate and timely temperature recording is helpful to society as a whole, we provide a novel channel of welfare gain through derivative contracts. The positive externality of better weather measurement could be enjoyed by the many businesses that rely on timely and accurate temperature measurements to make decisions. As an example, energy companies use both high and low frequency temperature data to plan energy production. An improvement in temperature accuracy will lead to better production planning by such companies. Indeed, NCDC has established a number of sector-specific user engagement programs that highlight the needs for timely and accurate data by a diverse set of industries such as energy, transportation, tourism, and construction.[6]

Finally, our study contributes to the corporate governance literature that focuses on the role of markets in disciplining corporate managers. Several studies analyze the role of markets as a monitor of top corporate managers. Gillan and Starks (1998) and Becht, Bolton, and Roell (2003) provide comprehensive surveys of the literature. We complement this literature by providing causal evidence from a non-corporate setting.

The rest of the paper is organized as follows. In Section 2 we describe the weather derivatives market in detail and highlight some key aspects of temperature measurement by the NWS. Section 3 describes the data and provides sample statistics. Section 4 provides the empirical design and results of the paper. Section 5 concludes.

2 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 $3 trillion of the U.S. GDP is associated with weather-sensitive industries (see Dutton, 2002 for details). Industries such as energy, construction, food processing, retail, and transportation are especially exposed to weather risk. Weather derivative products can provide insurance against weather related losses to these businesses. In addition, these products provide an alternative investment and

[6] see http://www.ncdc.noaa.gov/oa/userengagement/userengagement.html .

diversification opportunity to the financial investment community. While the need for insurance against weather conditions has been felt for a long time, it was only in 1999 that the first set of exchange traded weather contracts was listed on the Chicago Mercantile Exchange (CME). The exchange launched temperature based futures and options contracts on 10 U.S. cities within 13 months of September 1999. Subsequently, it 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.[7]

As of September 2005, approximately the middle point of our sample period, the total notional value of all CME traded weather contracts amounted to about $22 billion and an overwhelming majority of weather contracts are based on temperature. 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 2005-06 (WMRA Survey Report, 2006). Other major categories included contracts on rain, wind, and snow. Temperature related contracts insure the buyers from excessive heat or cold during the contract month. There are two types of contracts under this category: Heating Degree Days (HDD) and Cooling Degree Days (CDD) contracts. The buyer of a HDD contract receives payments for cold days defined as days with average temperature below 65^0F; conversely the buyer of a CDD contract receives payments for hot days defined as days with average temperature exceeding 65^0F. These contracts are written on observed temperature of a specific city for a specific period. As an illustration, consider a CDD option contract on Chicago for the month of August. The contract specifies a weather station in Chicago as the reference station for this trade. These weather stations are typically located near the underlying city's airport and are identified by their WBAN number. WBAN, an acronym for Weather-Bureau-Army-Navy, is a five-digit weather station number that uniquely identifies a measurement location. The Chicago contracts in our example are settled based on WBAN station number 94846 that is located at the O'Hare International Airport. Every day

[7]In addition to these 24 cities, CME also has snowfall contracts on Newark and the hurricane index on the Eastern US weather from Brownsville, Texas to Eastport, Maine. We do not include these two locations in our analysis since our focus is on city specific temperature related contracts.

in August the CDD contract compares the average of daily maximum and minimum temperature (T_{avg}) reported at this station with 65^0F and computes the cooling degree 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 tremendous economic implications for the market participants.

The final settlement of these contracts are based on the CDD or HDD index reported by MDA Information Systems, Inc. The settlement occurs on the second business day after the contract month.[8] MDA (formerly Earth Satellite Corporation, founded in 1969) is a private company and a leading provider of weather data to the weather trading industry. CME uses MDA's services to obtain temperatures based on NWS data for its trade settlements. MDA obtains weather data reported by the NWS and performs several quality control checks before transmitting it 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. For example, NWS occasionally reports missing temperature data for a weather station. The missing data can arise due to improper recording or other instrument malfunctions. In such cases, "MDA Federal first attempts to recover this data from alternative data sources, such as Climate Summary Reports, contacting the local NWS office or local media reports, as appropriate." (quoted from MDA's procedure manual). This is one direct example of increased outside scrutiny and visibility of the temperature numbers reported by the NWS.

2.1 Temperature Measurement

There are many government agencies that coordinate to meet the public's weather needs. The ultimate weather authority is the Department of Commerce (DOC), which is a Cabinet department of the federal government. Within the DOC, the National Oceanic and Atmospheric Administra-

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

tion (NOAA) is a bureau "focused on the condition of the oceans and the atmosphere." NOAA oversees 6 main offices, the 2 offices working with surface temperatures are the National Weather Service (NWS) and the National Environmental Satellite, Data and Information Service (NESDIS). The NWS handles most weather related government activities, including producing and disseminating temperature readings. The NESDIS manages and archives data collected by many government agencies. The National Climactic Data Center (NCDC) is an office within the NESDIS that archives and processes past weather records. In summary, the NWS and NOAA are the main agencies ensuring accurate on site measurement of temperatures throughout the United States, while the NCDC handles cleaning and storing past temperatures.

As mentioned earlier, weather derivative contracts are settled on the basis of temperature readings produced by the underlying WBAN stations. These stations are typically located at the city's main airport. A great degree of care is needed to obtain temperature with high accuracy even in a laboratory setting (e.g., see McGee (1988) for a detailed analysis of temperature measurement issues). These WBAN stations measure temperature in an outside environment, which can be even more difficult to measure with precision. A wide variety of factors affect accurate temperature measurement at a WBAN station. These factors can be broadly classified into three (non-exclusive) groups: (a) technological, (b) environmental, and (c) human. The technological factors relate to basic quality of the thermometer such as sensor's effectiveness, calibration errors, and self-heating of the instrument. Environmental factors relate to issues such as the location of the sensors and the effect of nearby electric disturbances, radiation, sunlight, and wind. The human factor captures the effect of manual intervention needed to measure accurate temperature. These interventions come in several forms such as active maintenance of the instrument, proper calibration, and minimizing the impact of environmental factors that can lead to inaccurate reports.

NOAA and NWS have detailed procedure manuals for collecting these readings in a timely and accurate manner. They also issue regular directives to their field offices on best practices in measuring temperature. These directives can be obtained from NOAA's website.[9] As an example, consider the NWS instruction 10-1302, dated June 21, 2010. It details out requirements and standards for NWS temperature and precipitation recordings.[10] It lays out procedures for proper installation, monitoring, and maintenance of these instruments. A few examples of these guidelines are: (a) the instrument must be placed at least 100 feet from any concrete or paved surface; (b) 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; (c) the instrument should not have any major obstruction (for example nearby buildings, trees, or fence) close-by that can affect its readings. Similarly, the NWS directive 10-1004 issued on February 17, 2011 provides a detailed set of instructions on monitoring of surface weather stations. These instructions point out the possible sources of error in temperature measurement and the NWS's attempts at training their staff to minimize these error rates. These guidelines also show the role of human interference in measuring these variables in an accurate manner.

In addition to CME and MDA, traders and financial parties regularly monitor these numbers and establish financial positions in this market based on their needs. Weather scientists have taken note of the increased attention paid to climate observations by the private sector in recent years (e.g., see Changnon and Changnon, 2010). As expected, the National Oceanic and Atmospheric Administration (NOAA), NWS and weather industry professionals have all recognized the need for better data quality from the WBAN stations. A number of initiatives such as joint conferences and exchange of ideas have taken place between these groups in light of the weather derivatives introduction. A workshop report in 2002 by the American Meteorological Society (Muranane et al. 2002) discusses the data needs of the private sector in the weather derivatives market.

We argue that the introduction of a weather derivative market attaches immediate and large economic importance to the NWS temperature numbers, which in turn results in tremendous scrutiny of these numbers by investors, media, and other related parties. Indeed, the NWS also

[9]http://www.nws.noaa.gov/directives/010/010.htm.

[10]See http://www.nws.noaa.gov/directives/sym/pd01013002curr.pdf

recognizes the need for better data collection exercise in light of the increased scrutiny by outside parties. In the Appendix, we provide an excerpt from an NWS directive to the field offices that highlights this aspect of monitoring. We also provide an excerpt from a meeting of NWS officials with weather industry representatives regarding the need for better data in the Appendix.

3 Data

We collect data from several sources and combine them together for our analysis. We first collect information on the launch dates of monthly derivative contracts on a city's temperature from the CME and press releases. For some cities, the CME introduced weekly and seasonal contracts at a later date as well. These contracts were introduced after the monthly contracts, hence we focus on the monthly contract dates. There are 24 weather stations with temperature derivative contracts as of the end of 2011. In addition, 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. Population data comes from the United States Census Bureau. We use the 2011 population estimates for metropolitan areas for this purpose.[11] We identify the WBAN number (i.e., the exact station number) of all derivative cities based on the contract specification. For the control cities, we use the weather station at the largest nearby 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 1. There have been four main waves of derivatives introduction: 1999-2000, 2003, 2005, and 2008. The list of derivative stations cover mostly large cities as well as a few smaller cities that are likely to have large economic interests tied to weather.

We obtain all weather data from MDA Information Systems, Inc. As mentioned earlier, MDA is a leading provider of weather data to weather traders as well as to the CME. We obtain two pieces of information for each weather station: (i) the raw temperature readings,

[11]http://www.census.gov/popest/data/metro/totals/2011/

and (ii) the cleaned temperature 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. We obtain data on the daily maximum and minimum temperature because the weather derivative contracts are settled based on the average of these two values. The raw temperature comes from METAR readings, which are standardized weather reports produced by Automated Surface Observing Systems stations. These stations are collectively operated by the Federal Aviation Administration, National Weather Service 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 WBAN station from the NWS METAR reports. The NWS stations produce hourly weather reports, 6-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 6-hour or hourly reports.

The second key measure is the 'cleaned' temperature value for every station-date pair. MDA uses a detailed five step process to clean the raw temperature values obtained from the government agencies. Through this process they ensure that the data is consistent with nearby reporting stations, and it conforms to meteorological consistency. They also take 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 has 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. Equally important, MDA checks all the raw temperatures for erroneous values by checking "the data against itself and against alternative data sources, such as hourly data, Climate Summary Reports, surrounding stations, and additional observations, as appropriate." MDA's meteorologists then examine the temperatures to ensure they are meteorologically consistent, i.e., they conform to basic consistency checks against other weather related variables. If temperatures are missing or erroneous, then new values are created using proprietary estimation techniques of the MDA. Using this detailed process, MDA arrives at a clean temperature measure that is used widely 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 use the difference between corrected and raw value as our key measure of measurement error in NWS temperature recording.

We also obtain data on temperature corrections by the NCDC. NCDC issues corrections to the NWS temperature numbers 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. Further information on preliminary and cleaned data can be obtained from NWS instruction manuals such as NWSI 10-1004 dated February 17, 2011 (NWS, 2011) and NWSPS 10-10 dated September 29, 2010 (NWS, 2010). NCDC restated numbers are extremely close to the MDA corrected values. Therefore, we use the difference between NWS raw numbers and MDA corrected values as the main variable in all our tests. We prefer the MDA based clean values because it alleviates the concern that the government agencies may be less inclined to restate their recordings after these contracts begin to trade.

3.1 Descriptive Statistics

Our sample covers all 49 stations from 2000 to 2011. We begin in 2000 purely because of data limitations. We are able to obtain good quality historical data on raw NWS temperature values only from the year 2000. Due to this limitation, we are not able to exploit the changes around the first set of derivatives introduction in 1999-2000. This leaves us with 14 stations for which we have data on both before and after the derivatives' introduction and we exploit the variations generated by these stations around the shock date in our empirical tests.

We take the number of days a given station reports erroneous or missing values as the main measure of temperature inaccuracy. These are the dates when the raw and cleaned values differ from each other. We aggregate this number at the yearly level and use the yearly count as the key measure of measurement error rate of a station in a year. We have 49 annual observations spread over 12 years (588 in total) in our sample. Some station-years show considerable error rate leading to skewness in the data. The Los Angeles weather station, for example, has very high levels of error rate across many years. In our empirical design we remove such station specific effects using station fixed effects. Further, we winsorize the data at 5% from both tails to ensure that our results are not driven by outlier observations. We also use log transformed error rate as an alternative measure of the dependent variable to alleviate concerns about outliers. Our results remain robust to either specification. Summary statistics are presented in Table 2. A representative station reports about 12 error days per year. There is considerable cross-sectional variation in the data as evident by the 90th (20 days) and 10th (5 days) percentiles of error days in the sample (see also Figure 1). In unreported results, we find that raw and final numbers remain the same for 96.69% of days. Of the remaining 3.31% days, 2.12% have a difference of 1^0F between the raw and cleaned data. The remaining 1.19% observations have considerably large discrepancies mostly ranging from 2-10^0F.

In addition to the main data on temperature recordings, we also obtain open interest and volume data for each station's temperature derivative contracts from the Chicago Mercantile Exchange. We use this information to analyze the relationship between derivative introduction and temperature accuracy across stations with high and low economic interest.

4 Empirical Design and Analysis

We estimate the effect of weather derivative introduction on the accuracy of temperature measurement in a difference-in-difference framework. We compare the measurement accuracy of the weather derivative station after the shock (i.e., after the introduction date) with the same station's accuracy before the shock to get the first margin of difference. We use the difference in the accuracy level of the non-shocked stations' measurement to remove the effect of any general changes in the NWS's weather measurement technology around the event date. The staggered nature of the shock allows us to remove the effect of any macro-economic factors and climatic variations that might occur at the same time as the derivatives introduction date. We implement this research design using the following regression model:

$$merr_{st} = \alpha_s + \beta \cdot derivative_{st} + year_t + \epsilon_{st} \tag{1}$$

$merr_{st}$ denotes measurement error at the WBAN station s in year t; α_s stands for station fixed effects; $year_t$ denotes the year fixed effects. $derivative_{st}$ takes a value of one for station-year observations after the introduction of derivatives, zero otherwise. The year of introduction is included in the post-introduction period. In this specification, 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-difference estimate of interest. The key identifying assumption is that the derivative's launch is not correlated with unobserved improvements, unrelated to our proposed channel, in the station's ability to measure the temperature. It is unlikely that the unobserved ability of the station officers change precisely at the 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.

We estimate this model using data on all 49 weather stations for the 2000-2011 period. As stated earlier, we have 14 stations which experience a derivative introduction after the year 2000. Of the remaining, 10 stations had derivatives trading throughout the entire sample period, whereas 25 of them do not have weather derivative contracts at all. Thus our sample provides a balanced mix of stations along the key dimension of interest and allows us to separate the effect of derivative introduction on measurement accuracy by estimating the changes around the introduction date for the shocked station.

As a prelude to our regression analysis, we provide the average number of error days for the shocked stations before and after the shock and compare that to the corresponding averages of the control firms. We take the average number of error days for all control stations during the given calender year for this exercise. We compute the average error days across the two groups for the periods before and after the shock and plot them in Figure 2. There is a drop in the number of error days from before to after the shock for both groups. However, the shocked stations experience a drop of 1.54 as compared to the control station's drop of 0.68. 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 3. Models 1 and 2 use the raw number of error days as the dependent variable, whereas Models 3 and 4 use its log transformed values. Model 1 presents the results without year fixed effects. We obtain a coefficient of -2.24 on the *derivative* variable, indicating a decline of about 2.25 days in the annual error rate. The effect is statistically significant at the 1% level. In Model 2 we include the year fixed effects to remove the effect of any secular improvement in weather measurement technology over time or the effect of other macroeconomic and climatic changes that might effect the measurement error of all stations. We obtain a coefficient estimate of -1.11 that is significant at the 5% level. This estimate translates into a decline of about 10% in the error rate of the median station after the introduction of the derivatives. Models 3 and 4 obtain similar results and ensure that our results are not driven by outliers. These baseline results establish the effect of derivatives introduction on measurement accuracy: NWS reported raw numbers become systematically closer to the cleaned values once there is a direct financial market interest tied to these numbers.

As a robustness exercise, we check for and rule out the presence of any pre-existing trends in the error days of the shocked stations. We compute the change in error days from three years before to the year before the shock date. The shocked stations experience an average change of $+0.64$ during this period. In contrast, they show a decline of -3.93 from a year before the shock till three years after it. This shows that there is no secular trend of improvement in measurement accuracy before the launch. The improvement, therefore, is likely caused by the introduction of financial markets.

We also analyze these effects at the cohort level as a further robustness exercise. For this test, we consider one derivative introduction cohort at a time. We take all the shocked stations for the given cohort and include data from 2000 to three years after the introduction year in the sample. We limit the sample to three years post-derivative introduction to estimate our main effects in the immediate aftermath of the launch. As an example, for the 2003 cohort, we include data from 2000 to 2006 for all the stations that launched derivative contracts in 2003 (Kansas City, Houston, Boston, Minneapolis, and Sacramento) in the treatment group and include all the non-derivative stations in the control group. Results are provided in Table 4. We obtain a negative coefficient on the *derivative* variable for all three cohorts, but the effects are significant only for the first two of the three cohorts. The strongest effect comes from the 2003 cohort and monotonically declines for the next two. These results show that immediately after the launch of the weather derivatives market, there has been a noticeable change in the accuracy rate of the affected stations and the effect is more pronounced for the earlier cohorts. Over time best measurement practices are likely to be shared by all NWS stations, making the effect of derivative introduction less impactful. More important, we find that earlier cohort stations have much higher open interest in the underlying derivatives as compared to the later ones. This could be partly driven by the fact that CME chooses to introduce derivative contracts in cities with high demand first. Thus, our evidence that the effects are more pronounced in the earlier cohorts is consistent with the idea that higher economic interests leads to higher visibility and better monitoring effects. We explore this channel more formally in the next section.

4.1 Economic Interests

We show that the introduction of temperature related financial contracts results in better measurement outcomes by the NWS. We argue that these effects arise due to increased economic interests and the resulting scrutiny of these measures by the markets after the derivatives launch. In this section we provide some direct evidence to support this view. Our tests are designed to exploit the cross-sectional variation in the level of economic interests that in turn is likely to be correlated with the level of scrutiny and reputational pressure on the weather measurement stations.

In the first test, we assess the effect of launch on measurement errors across stations with varying degree of open interest in the weather derivatives market. We obtain data on monthly open interest at these stations over the 2007-2012 period.[12] We compute average monthly open interest over this period for each station in our sample and create an indicator variable based on this number. Out of the 24 stations, 16 stations have active trading in this market. We classify a station as a high economic interest market if it falls in the top 16 stations, and as a low economic interest market if it is in the bottom 8. Out of the 14 derivative stations that have their launch date after 2000, there are 6 in the high and 8 in the low open interest group.

We break the *derivative* variable into two groups based on whether the station falls in high or low open interest group. *High Open Int.* takes a value of one for high open interest derivatives stations after the derivative launch, and zero otherwise. *Low Open Int.* variable is computed in the similar manner. We re-estimate the main regression model with these two variables as the regressors and provide the results in Table 5. While both groups show some improvement post-derivatives launch, it is mainly coming from the high open interest group. The result is consistent with our argument that markets work as a disciplining device on the agents responsible for measuring the reference temperature. As mentioned earlier, the low open interest stations are mostly concentrated in the later cohort. Thus theses results are based on roughly the same sources of variation as the results produced in the cohort-by-cohort analysis in the previous section.

In the second test, we estimate the effect of the derivatives market on temperature accuracy across high and low population cities. Based on each city's population, we create an indicator variable for the top 25 cities. We create two variables based on derivative trading and the city's population: *High Pop* and *Low Pop*. The first variable equals one for derivative stations that fall in top 25 population cities, and zero otherwise. Similarly, the second variable equals one for derivative stations that fall outside the top 25 population cities, and zero otherwise. Out of

[12] We have been able to obtain the open interest and volume data only for this period as of now. We are in the process of collecting data for the earlier periods.

the 24 derivative trading stations, 14 are in the high population group. Among the derivative trading stations that launched their product during our sample period, there are 8 stations in the high population group. Results are provided in Table 6. We find that the improvements are concentrated in the subset of high population cities. These cities are likely to have higher economic interests tied to derivative contracts and for these cities the reputation loss of mistakes are likely to be higher. Our results are consistent with the notion that measurement outcomes improve due to such concerns.

As expected, our results are strongest among the subset of stations that qualify as high economic interest based on the intersection of the two measures discussed above. Weather stations that fall in high open interest category as well as high population group show the most improvement in measurement accuracy. For brevity, we do not tabulate these results.

4.2 Channel of Improvement

We show that the introduction of financial markets improves the actions of NWS by bringing more visibility and scrutiny of the reported temperature numbers. We now focus on the sources of improvement. In particular, there are two possible, not mutually exclusive, channels of improvement. First, the NWS might install better thermometers/sensors at these stations precisely at the time when derivatives start trading. We call this the *technology* channel. Second, the NWS might start putting in more effort to better capture the temperature data in an accurate manner. We call this the *human* channel. While the net effect of both these channels remains the same, i.e., an improvement in the measurement of weather, our focus is more on the second channel. Said differently, we want to investigate the disciplining effect of market purely on account of higher effort put in by the government officials. These improvements can come through better maintenance and monitoring of the weather stations to minimize the erroneous reporting.

We separate the two channels by exploiting an important cross-sectional variation in trading activities across months in the weather derivatives market. The end-users of this market typically comprise businesses in sectors such as utilities, transportation, retail, and food products. A majority of their hedging demands arise in the months of extreme heat or cold. Not surprising, an overwhelming majority of these contracts are based in peak summer and peak winter months. This leaves the months of April and October as the least traded months on the exchange. We separately estimate the basic regression model by first using only April and October and then using observations from all months except April and October. The key idea is to assess the improvement in measurement efforts keeping the underlying measurement technology the same.

We aggregate all the error days in April and October for the first analysis and similarly the error days in the remaining months for the second analysis. Results are provided in Table 7. We find no improvement in October and April, whereas there is significant improvement in the active trading months. As a robustness we separate the sample into two groups by clubbing ± 1 month around April and October in one group, and the rest in another. We find negative coefficients (unreported) for both groups, but the peak months' regression coefficient is roughly double the off-peak months' regression coefficient. Again, the result is significant only for the peak months. Thus even at the same station, the improvement comes from months with active trading. These are the months where the pressure and monitoring from the outside market is likely to be the highest. In these months, the frequency of follow ups with the NWS stations and analysis of the weather data by the trading professionals is expected to be higher than the remaining months. Our result supports the view that financial markets induce higher effort by the stations in measuring the temperature accurately.

4.3 Estimation with NCDC Measure

All our results so far have been 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 clean or restated data produced by an affiliated government agency of the NWS, namely the NCDC. The NCDC is responsible for producing the final data after removing data collection errors at the station. These cleaned data become NOAA's official data and are widely used in meteorological studies.

We begin with all 49 stations in our sample. However, NCDC did not produce corrected values for 8 stations in December, 2001. In addition, the agency did not produce cleaned values for two control stations (San Jose and Riverside) during our sample period. Hence, we lose two control stations and one observation month for this part of the study. We re-estimate our main regression models based on the NCDC data and report the results in Table 8. As can be seen from the Table, the results are almost identical to the ones reported using MDA values. These results provide confidence in our measure of temperature accuracy since data from both these parties – MDA, a third-party private company, and NCDC, an affiliated government agency – produce similar results. Their broad agreement with the cleaned or correct temperature value alleviates the concern that we might have a bad measure of temperature accuracy. Indeed, on the set of overlapping observations, common to both MDA and NCDC, we find that they agree on the correct temperature values in almost all cases. There are only 4 instances out of over 200,000 daily observations where there is a disagreement between the two agencies about the correct temperature value. Therefore, it is not surprising that we get almost identical results using either one of these measures.

5 Conclusion

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 billions of dollars of contracts in the private market. Thus there is an increased interest 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.

Our results highlight an important role of financial markets. They can work as a disciplining device even in the absence of explicit incentives and monitoring mechanisms that are present in the corporate settings. Here, the numbers reported by a government agency become more accurate after the markets open up. To the extent that we care about accurate measurement of these numbers, there is a positive externality that comes from the financial markets. Indeed there are several industries, most notably the energy sector, that directly benefit from high frequency accurate data. We provide some evidence in support of this claim in the Appendix. Overall, our study provides one of the first empirical estimates of the impact of financial innovations on the real outcomes produced by parties that are not directly affected by the payoffs from the contract.

Appendix: Some Descriptive Evidence & Supporting Claims

In this appendix we produce pieces of evidence collected from several sources such as the NWS directives, NOAA, weather trading industries, and atmospheric science journals that are relevant to our study. We present some key facts and opinions from these sources as well as our summary of the material below:

1. NWS directives on data collection exercise: NWS issues directives to its regional offices and weather stations on a regular basis on a range of issues including data quality control and assurance standards. Some of these directives highlight the need for more accurate and consistent data in light of increased outside scrutiny. We provide an example from the NWS's directive (number NWSI 10-1305) issued on April 28, 2008:

   "The NWS has the responsibility of collecting and providing weather and climate observation data. However, the methods for the collection, quality control, and delivery of these data vary from office to office. Many of the data quality initiatives between the NWS and NCDC have been uncoordinated. Even with the NWS itself such activities vary greatly between field offices. This situation must change in the interest of efficiency, data record integrity and public use.

   Today, with the ever increasing use of observational data by the research community, the media, private industry, and the general public it is of the utmost importance to accurately and consistently apply QC/QA at all field offices. In order to ensure the highest quality data and data products within Central Region, the QC/QA methods discussed in this supplement are highly recommended at each WFO."

   Note: Emphasis added by the authors. QC/QA stand for quality control and quality assurance in the above quote.

2. NOAA's information on preliminary (i.e., raw) versus official (i.e., clean) data: Below we provide some examples of Frequently Asked Questions and their Answers from the NOAA's web-site about the raw versus cleaned data.

   "Are the data in NOWData considered 'official' for legal and other such purposes?

   No. NOWData provides up-to-date information based on archived AND preliminary data holdings by NOAA. For official data, you should contact NOAAs National Climatic Data Center or the Regional Climate Centers. NCDC provide official certification for data being used in U.S. courts."

   "I noticed that the most recent data does not match data that I found of the NCDC web site. Why is that?

   Preliminary data can be different from NCDC official data for a number of reasons related to quality assurance and processing schedules, as well as synchronization of the NCDC and ACIS databases. Ultimately, when processing is completed, the two data files will match."

   Note: NOWData stands for NOAA Online Weather Data, which comes from METAR readings, which is also our source of initial data recording.

3. A summary of the meetings between NWS and weather industry representatives: There have been quite a few meetings between the NWS officials and the weather derivative

professionals regarding the weather derivatives market. The weather industry has often expressed its need for better quality data from the weather station. Here is an excerpt from a meeting of NOAA staff with the representatives of weather derivative industry during the very early stages of this market (meeting dated March 12, 1998)[13]. This meeting occurred before the launch of official CME contracts.

"Data issues, both short and long-term, pertaining to these contracts were the immediate reason for this meeting. On their own initiative, industry participants have chosen to use daily temperature data from the National Weather Service to calculate their cumulative degree day indices upon which the contracts are based and which will be used to settle the contracts.

One concern they had was regarding the difference between preliminary and official data. NOAA indicated that the preliminary data are usually quite close to the official historical data, which are published with a lag of two to three months. With this understanding, the firms said they felt more comfortable using the preliminary data for initial settlement of the contracts, subject to a "true-up" to the official data several months later.

A second interest was that there be one set of tailored data for common reference. This could reduce disputes that might arise from different sources for the weather data."

4. Meridian Environmental Technology is a company specializing in atmospheric information and technology (amongst other things). Their website provides evidence of private enterprise's need for accurate weather information:

   "Power production planning requires accurate and reliable weather information. Meridian has been providing historical and forecasted site-specific weather information to the agriculture, transportation, and utilities industries for years. Whether you are needing hourly, daily, weekly or longer information, Meridian can help you!

   We understand your needs for forecasted power production and the high penalties a wrong estimate can cost..."[14]

5. NOAA's NCDC Sectoral Engagement Fact Sheets[15] document industries that depend on quality weather information from NOAA. NOAA lists Agriculture, Civil Infrastructure, Coastal Hazards, Energy, Health, Insurance, Litigation, Marine and Coastal Ecosystems, National Security, Tourism, Transportation and Water Resources as industries sensitive to the climate. Not all of these industries will be directly affected by inaccurate temperature measurements, but some are. For example, in the Energy Fact Sheet NOAA writes about how companies are:

   "Using temperature information to aid in the assessment of equipment requirements for heavy power line loads during extremely hot weather."

   These loads will be determined by weather measurements produced by the government. If the numbers are incorrect, energy companies may use the incorrect amount or type of equipment.

[13] See the full document at http://www.srh.noaa.gov/topics/attach/html/ssd98-14.htm
[14] http://www.meridian-enviro.com/pages.pl?pg=usf
[15] http://www.ncdc.noaa.gov/oa/userengagement/userengagement.html

References

1. Becht, M., P. Bolton, and A. Roell, 2003, Corporate Governance and Control, Handbook of The Economic of Finance, Volume 1A, Elsevier North Holland.

2. Changnon, D., and S. A. Changnon, 2010, Major Growth in Some Business-Related Users of Climate Information, Journal of Applied Meteorology and Climatology 49, 325-331.

3. Diamond, D., 1989, Reputation Acquisition in Debt Markets, Journal of Political Economy 97, 828-862.

4. Dutton, J. A., 2002, Opportunities and Priorities in a new Era for Weather and Climate Services, Bulletin of the American Meteorological Society 83(9), 1303-1311.

5. Gillan, S., and L. Starks, 1998, A Survey of Shareholder Activism: Motivation and Empirical Evidence, Contemporary Finance Digest 2, 10-34.

6. Holmstrom, B., and J. Tirole, 1993, Market Liquidity and Performance Monitoring, Journal of Political Economy 101, 678-709.

7. Khan, C., and A. Winton, 1998, Ownership Structure, Speculation, and Shareholder Intervention, Journal of Finance 53, 99-129.

8. Levitt, S., and J. List, 2011, Was There really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments, American Economic Journal: Applied Economics 3, 224-238.

9. McGee, T. D., 1988, Principles and Methods of Temperature Measurement, John Wiley & Sons, Inc.

10. Murnane, R. J., M. Crowe, A. Eustis, S. Howard, J. Koepsell, R. Lefflet, and R. Livezey, 2002, The Weather Risk Management Industry's Climate Forecast and Data Needs: A Workshop Report, Bulletin of the American Meteorological Society 83(8), 1193-1198.

11. MDA Federal, Procedural Manual, Chicago Mercantile Exchange. http://www.cmegroup.com/trading/weather/files/procedure-manual.pdf

12. National Weather Service, 2008, Data Quality Control and Assurance Standards, National Weather Service Instruction 10-1305, NWS, Silver Spring, MD.

13. National Weather Service, 2010, Climate Data Services, National Weather Service Instruction 10-1003, NWS, Silver Spring, MD.

14. National Weather Service, 2010, Climate Data Services, National Weather Service Instruction 10-1302, NWS, Silver Spring, MD.

15. National Weather Service, 2011, Climate Data Services, National Weather Service Instruction 10-1004, NWS, Silver Spring, MD.

16. National Weather Service, 2011, Climate Records, National Weather Service Instruction NWSPD 10-10, NWS, Silver Spring, MD.

17. National Oceanic and Atmospheric Administration, 2010, NOAAs National Climatic Data Center Sectoral Engagement Fact Sheet Energy, NOAA, Silver Spring, MD.

18. Perez-Gonzalez, F. and H. Yun, forthcoming, Risk Management and Firm Value: Evidence from Weather Derivatives.

19. PricewaterhouseCoopers, Weather Risk Management Association Survey, 2006.

20. Shleifer, A., and R. W. Vishny, 1986, Large Shareholders and Corporate Control, Journal of Political Economy 94, 461-488.

21. Tufano, P., 2003, Financial Innovation, Handbook of The Economic of Finance, Volume 1A, Elsevier North Holland.

22. Wilson, J. Q., 1989, Bureaucracy: What Government Agencies Do and Why?, Basic Books, Inc.

Table 1: Weather Stations

This table presents information on the weather stations in our sample. Station Name is the name of the weather station. WBAN is a station identifier. Introduction date is the first trading day of the first temperature contract on that location. Major City is the nearest city. Pop. Rank is the 2011 metropolitan population rank amongst United States metropolitan areas. OI Rank is the location's rank amongst the 24 derivative locations based on total open interest between 2007 and 2012.

Station Name	**WBAN**	**Introduction Date**	**Major City**	**Pop. Rank**	**OI Rank**
Weather Derivative Stations					
New York LaGuardia Airport	14732	September 22, 1999	New York, NY	1	1
Atlanta Hartsfield International Airport	13874	September 22, 1999	Atlanta, GA	9	2
Chicago O'Hare International Airport	94846	September 22, 1999	Chicago, IL	3	3
Cincinnati Northern Kentucky Airport	93814	September 22, 1999	Cincinnati, OH	27	5
Dallas-Fort Worth International Airport	3927	September 30, 2000	Dallas, TX	4	4
Des Moines International Airport	14933	September 30, 2000	Des Moines, IA	88	6
Philadelphia International Airport	13739	September 30, 2000	Philadelphia, PA	6	7
Las Vegas McCarran International Airport	23169	September 30, 2000	Las Vegas, NV	30	12
Portland International Airport	24229	September 30, 2000	Portland, OR	23	14
Tucson International Airport	23160	September 30, 2000	Tucscon, AZ	52	15
Minneapolis-St. Paul International Airport	14922	September 26, 2003	Minneapolis, MN	16	8
Kansas City International Airport	3947	September 26, 2003	Kansas City, MO	29	9
Boston Logan International Airport	14739	September 26, 2003	Boston, MA	10	10
Houston Bush Intercontinental Airport	12960	September 26, 2003	Houston, TX	5	11
Sacramento Executive Airport	23232	September 26, 2003	Sacramento, CA	25	13
Baltimore-Washington International Airport	93721	June 20, 2005	Baltimore, MD	20	16
Salt Lake City International Airport	24127	June 20, 2005	Salt Lake City, UT	48	18
Detroit Metropolitan Airport	94847	June 20, 2005	Detroit, MI	13	20
Raleigh Durham International Airport	13722	May 19, 2008	Raleigh Durham, NC	47	-
Little Rock Adams Field	13963	May 19, 2008	Little Rock, AR	72	17
Washington Reagan National Airport	13743	May 19, 2008	Washington D.C.	7	19
Colorado Springs Municipal Airport	93037	May 19, 2008	Colorado Springs, CO	81	21
Los Angeles Downtown USC Campus	93134	May 19, 2008	Los Angeles, CA	2	22
Jacksonville International Airport	13889	May 19, 2008	Jacksonville, FL	40	23

Table 1 continued

Station Name	**WBAN**	**Introduction Date**	**Major City**	**Pop. Rank**	**OI Rank**
Control Stations					
Austin-Bergstrom International	13904	-	Austin, TX	34	-
Nashville International Airport	13897	-	Nashville, TN	37	-
Cleveland Hopkins International Airport	14820	-	Cleveland, OH	28	-
Charlotte Douglas International Airport	13881	-	Charlotte, NC	33	-
Columbus Port Columbus International Airport	14821	-	Columbus, OH	32	-
Denver International Airport	03017	-	Denver, CO	21	-
Indianapolis International Airport	93819	-	Indianapolis, IN	35	-
Orlando International Airport	12815	-	Orlando, FL	26	-
Memphis International Airport	13893	-	Memphis, TN	41	-
Miami International Airport	12839	-	Miami, FL	8	-
Milwaukee Mitchell International Airport	14839	-	Milwaukee, WI	39	-
Oklahoma City Will Rogers World Airport	13967	-	Oklahoma City, OK	43	-
Norfolk International Airport	13737	-	Norfolk, VA	36	-
Phoenix Sky Harbor International Airport	23183	-	Phoenix, AZ	14	-
Pittsburgh International Airport	94823	-	Pittsburgh, PA	22	-
Providence T F Green State Airport	14765	-	Providence, RI	38	-
Riverside Municipal Airport	03171	-	Riverside, CA	12	-
San Diego Lindbergh Field	23188	-	San Diego, CA	17	-
San Antonio International Airport	12921	-	San Antonio, TX	24	-
Louisville Standiford Field	93821	-	Louisville, KY	42	-
Seattle Seattle-Tacoma International Airport	24233	-	Seattle, WA	15	-
San Francisco International Airport	23234	-	San Francisco, CA	11	-
San Jose International Airport	23293	-	San Jose, CA	31	-
St Louis Lambert International Airport	13994	-	St. Louis, MO	19	-
Tampa International Airport	12842	-	Tampa, FL	18	-

Table 2: 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 25 control weather stations.

Year	**N**	**Mean**	**Median**	**SD**	**10th**	**90th**
2000	49	11.428571	11	5.2717802	4	19
2001	49	11.122449	10	5.2544927	5	19
2002	49	15.387755	16	5.0366007	8	22
2003	49	14.326531	15	6.2094946	5	22
2004	49	13.428571	13	5.5976185	6	22
2005	49	11.44898	11	4.7392915	5	17
2006	49	10.571429	9	4.8175028	4	17
2007	49	12.693878	13	4.9379655	7	20
2008	49	12.428571	12	4.495368	5	19
2009	49	10.632653	10	3.8928025	4	15
2010	49	10.489796	10	4.3642604	4	16
2011	49	10.714286	10	4.1932485	5	17
All	588	12.056122	12	5.1321793	5	20

Table 3: The Effect of CME Derivative Introduction on Weather Station Errors

The dependent variable in columns 1 and 2 is the total number of weather station errors for each station-year observation. The dependent variable in columns 3 and 4 is the log of the total number of weather station errors for each station-year observation. All regressions include station fixed effects. Columns 2 and 4 include year fixed effects. Standard errors are clustered by weather station.

VARIABLES	(1) Total Errors	(2) Total Errors	(3) Log(Total Errors)	(4) Log(Total Errors)
Derivative	-2.245***	-1.109**	-0.167***	-0.102**
	(0.421)	(0.542)	(0.0306)	(0.0415)
Observations	588	588	588	588
R-squared	0.451	0.533	0.470	0.536
Year dummies	No	Yes	No	Yes
Station Dummies	Yes	Yes	Yes	Yes

Clustered Standard Errors in Parentheses
*** p<0.01 ** p<0.05 * p<0.1

Table 4: The Effect of CME Derivative Introduction on Weather Station Errors by Introduction Cohort

The dependent variable is the total number of weather station errors for each station-year observation. The regression in Column 1 includes weather stations with CME derivative introduction in year 2003 and all control stations for the years 2000-2006. The regression in Column 2 includes weather stations with CME derivative introduction in year 2005 and all control stations for the years 2000-2008. The regression in Column 3 includes weather stations with CME derivative introduction in year 2008 and all control stations for the years 2000-2011. All regressions include station and year fixed effects. Standard errors are clustered by weather station.

	(1)	(2)	(3)
VARIABLES	Total Errors	Total Errors	Total Errors
2003 shock	-2.613**		
	(1.109)		
2005 shock		-2.325**	
		(0.862)	
2008 shock			-0.600
			(0.807)
Observations	235	261	372
R-squared	0.606	0.591	0.540
Year dummies	Yes	Yes	Yes
Station Dummies	Yes	Yes	Yes

Clustered Standard Errors in Parentheses
*** $p<0.01$ ** $p<0.05$ * $p<0.1$

Table 5: Open Interest and The Effect of CME Derivative Introductions on Weather Station Errors

The dependent variable in column 1 is the total number of weather station errors for each station-year observation. The dependent variable in column 2 is the log of the total number of weather station errors for each station-year observation. Low Open Int. is an indicator for a derivative being traded on that station in that year interacted with an indicator for the station having total open interest in the bottom 8 of 24 stations over the years 2007-2012. High Open Int. is an indicator for a derivative being traded on that station in that year interacted with an indicator for the station having total open interest in the top 16 of 24 stations over the years 2007-2012 . All regressions include year and station fixed effects. Standard errors are clustered by weather station.

VARIABLES	(1) Total Errors	(2) Log(Total Errors)
Low Open Int.	-0.911	-0.0747*
	(0.661)	(0.0394)
High Open Int.	-1.337*	-0.134**
	(0.712)	(0.0633)
Observations	588	588
R-squared	0.533	0.536
Year dummies	Yes	Yes
Station Dummies	Yes	Yes

Clustered Standard Errors in Parentheses
*** $p<0.01$ ** $p<0.05$ * $p<0.1$

Table 6: Population and The Effect of CME Derivative Introductions on Weather Station Errors

The dependent variable in column 1 is the total number of weather station errors for each station-year observation. The dependent variable in column 2 is the logarithm of the total number of weather station errors for each station-year observation. High Pop is an interaction between an indicator for a station being located near a city with a top 25 population and a weather derivative being traded on that station in that year. Low Pop is an interaction between an indicator for a station being located near a city with a population outside of the top 25 and a weather derivative being traded on that station in that year. All regressions include station and year fixed effects. Standard errors are clustered by weather station.

	(1)	(2)
VARIABLES	Total Errors	Log(Total Errors)
High Pop	-2.164***	-0.168***
	(0.503)	(0.0469)
Low Pop	0.233	-0.0186
	(0.493)	(0.0395)
Observations	588	588
R-squared	0.536	0.537
Year dummies	Yes	Yes
Station Dummies	Yes	Yes

Clustered Standard Errors in Parentheses

*** p<0.01 ** p<0.05 * p<0.1

Table 7: Active Months and The Effect of CME Derivative Introductions on Weather Station Errors

The dependent variable in both regressions is the total number of weather station errors for each station-year observation. The regression in column 1 **excludes** all months except April and October. The regression in column 2 **includes** all months except April and October. Derivative is an indicator equal to 1 if a derivative is traded on the station in that year. All regressions include year and station fixed effects. Standard errors are clustered by weather station.

VARIABLES	(1) Total Errors	(2) Total Errors
Derivative	0.0913	-1.283**
	(0.252)	(0.554)
Observations	588	588
R-squared	0.216	0.515
Year dummies	Yes	Yes
Station Dummies	Yes	Yes

Clustered Standard Errors in Parentheses
*** p<0.01 ** p<0.05 * p<0.1

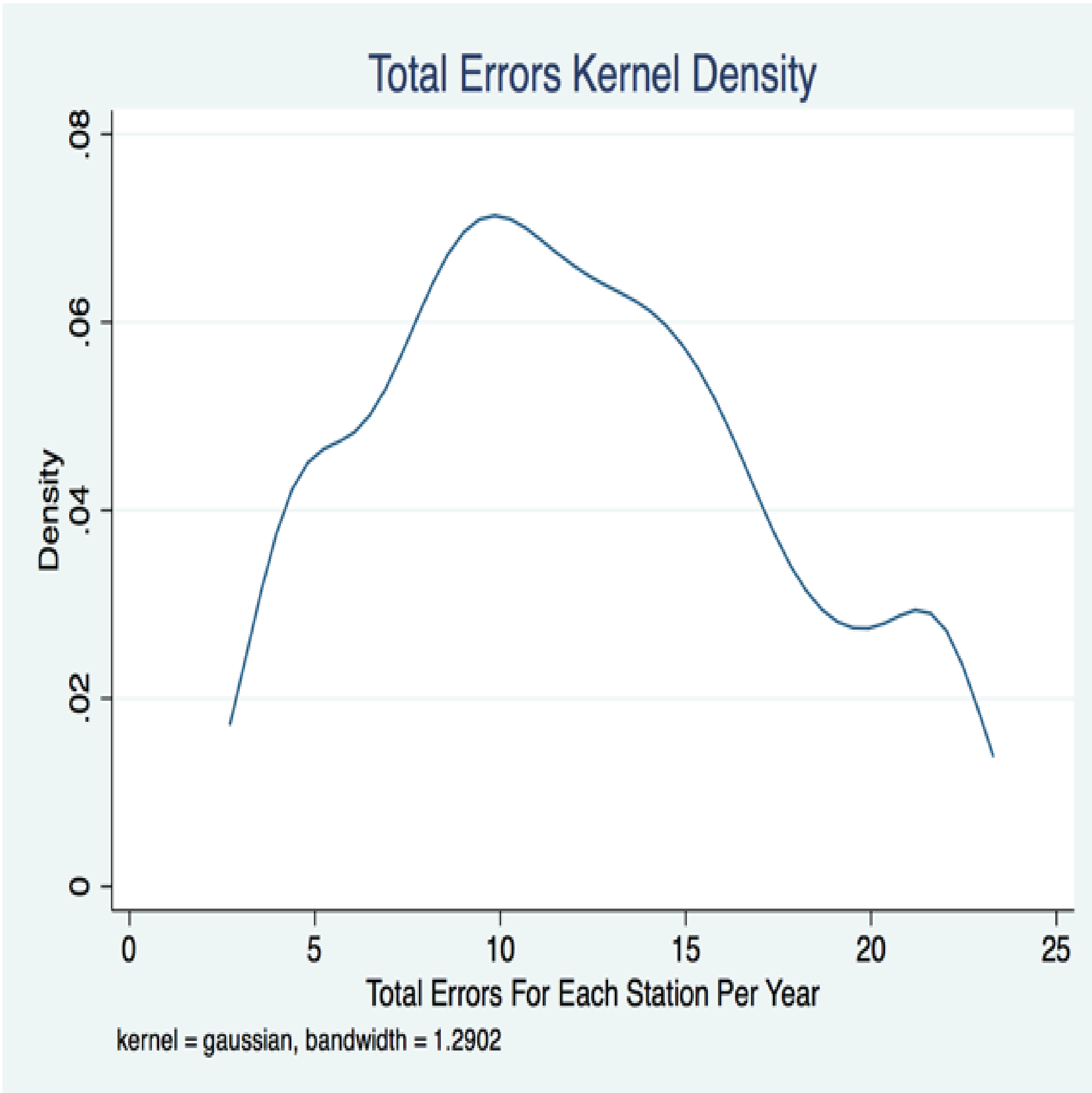
Table 8: The Effect of CME Derivative Introduction on Weather Station Errors Using NCDC Cleaned Values

The dependent variable in columns 1-2 & 5-7 is the total number of weather station errors for each station-year observation. The dependent variable in columns 3 and 4 is the log of the total number of weather station errors for each station-year observation. Derivative, 2003 shock, 2005 shock and 2008 shock are all a dummy equal to 1 if a weather derivative is traded on that location in that year. The regression in Column 5 includes only the control stations and those stations that received a derivative in 2003 for the years 2000-2006. The regression in Column 6 includes only the control stations and those stations that received a derivative in 2005 for the years 2000-2008. The regression in Column 7 includes only the control stations and those stations that received a derivative in 2008 for the years 2000-2011. The San Jose and Riverside stations are not included in these regressions. All regressions include station fixed effects. Columns 2 & 4-7 include year fixed effects. Standard errors are clustered by weather station.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Total Errors	Total Errors	Log(Total Errors)	Log(Total Errors)	Total Errors	Total Errors	Total Errors
Derivative	-2.169***	-0.918*	-0.175***	-0.0963*			
	(0.417)	(0.545)	(0.0319)	(0.0498)			
2003 shock					-2.370**		
					(1.132)		
2005 shock						-2.130**	
						(0.889)	
2008 shock							-0.382
							(0.850)
Observations	564	564	564	564	221	243	348
R-squared	0.466	0.554	0.480	0.542	0.628	0.617	0.567
Year dummies	No	Yes	No	Yes	Yes	Yes	Yes
Station Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes

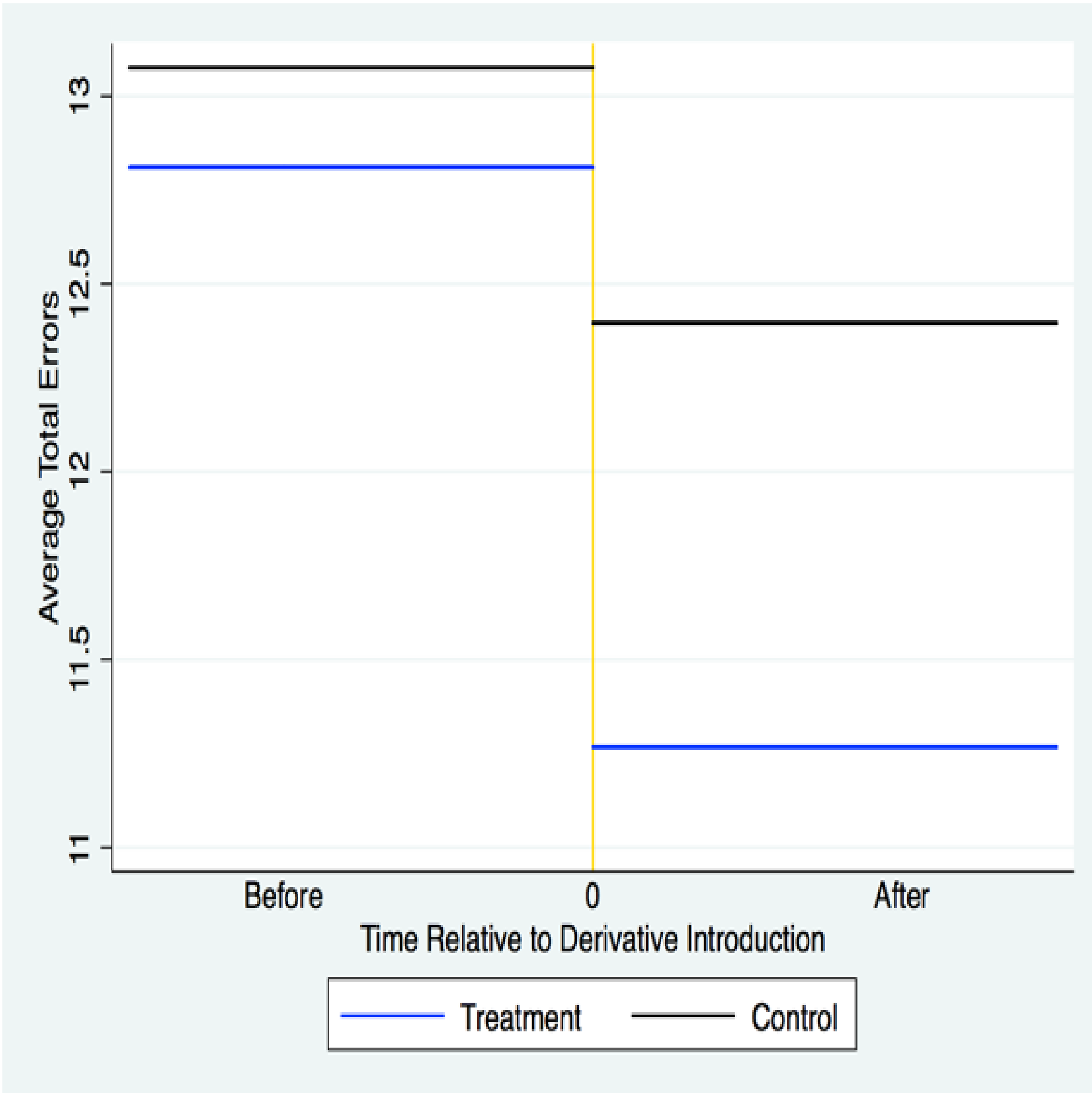
Clustered Standard Errors in Parentheses
*** p<0.01 ** p<0.05 * p<0.1

Figure 1: Weather Station-Year Total Errors Distribution



This figure shows the kernel density of the total errors each year for each weather station in our sample. We use the Gaussian kernel and a bandwidth that minimizes the mean integrated squared error assuming the data were Gaussian.

Figure 2: Average Yearly Errors Pre- and Post- Introduction)



This figure graphs the average errors for the treatment and control groups before and after weather derivative introduction. The before period is the 3 years Before introduction and the After period is the year of introduction plus the 3 years afterwards. The treatment group consists of the 14 weather stations that experienced a weather derivative introduction during our sample period and the control group consists of the 25 stations that never experienced a weather derivative introduction during our sample period.