North American Journal of Economics and Finance 80 (2025) 102494

Contents lists available at ScienceDirect

North American Journal of Economics and Finance

journal homepage: www.elsevier.com/locate/najef

Can extreme weather forecasts lead to a risk premium? Evidence of a non-linear response in U.S. natural gas futures ☆

Manou Monteux [a,b], Maria Cristina Arcuri [c,d,*], Gino Gandolfi [c,d], Stefano Caselli [e,f]

[a] Subject Matter Expert in Financial Institutions at Department of Economics and Management, University of Parma, Via J.F. Kennedy 6, 43125 Parma, Italy
[b] Genesis Machine Learning Solutions LTD, Netherwood Road 60, W14 0BG London, United Kingdom
[c] Department of Economics and Management, University of Parma, Via J.F. Kennedy 6, 43125 Parma, Italy
[d] Knowledge Group Banking & Insurance, SDA Bocconi School of Management, Italy
[e] Bocconi University, Dean of SDA Bocconi School of Management and Member of the Board of Directors of SDA Bocconi School of Management, Via Sarfatti 25, 20136 Milan, Italy
[f] Algebris Chair in "Long-Term Investment and Absolute Return", Algebris Investments, Corso Vittorio Emanuele II, 1, 20122 Milano, Italy

ARTICLE INFO

JEL classification:
G11
G13
G14
Q47
Q54
Q58

Keywords:
Alternative data
Risk premium
Weather risk
Natural gas futures
Market efficiency
Event study
Climate change

ABSTRACT

Using data on hourly frequency observed temperature and daily forecasted temperatures across major U.S. metropolitan areas over a 30-year period, we analyze the relationship between the daily returns of the NYMEX Henry Hub Natural Gas futures and U.S. weather fluctuations. We propose the existence of a novel risk premium linked to extreme weather forecasts for U.S. Natural Gas futures, which outperforms the S&P500 index on an absolute and risk-adjusted basis over a 30-year period. Our findings contribute to opening a new perspective on the non-linear interplay between weather and financial markets emphasizing the importance of these factors in financial risk management and in the context of climate change.

1. Introduction

Engle et al. (2020) state that Earth's climate is changing, but uncertainty around the trajectory and the economic consequences of climate change is rife and as a result, investors around the world are in the search for products that will help them to hedge against climate risk. While most economic studies on climate change focus on long term effects of climate change risk, which is a non-diversifiable risk (Engle et al., 2020), in this work we highlight a novel *extreme weather forecast* risk premium affecting the Natural Gas futures (NG) market. Although this novel risk premium predates today's climate change evidences, it implies a market mechanism where a higher level of extreme weather events, driven by climate change (Cohen et al., 2018; Cohen et al., 2021; Francis et al., 2012;

☆ This article is part of a special issue entitled: 'FinEc of Climate' published in North American Journal of Economics and Finance.

* Corresponding author at: Department of Economics and Management, University of Parma, Via J.F. Kennedy 6, 43125 Parma, Italy.
E-mail addresses: mariacristina.arcuri@unipr.it (M.C. Arcuri), gino.gandolfi@unipr.it (G. Gandolfi), stefano.caselli@unibocconi.it (S. Caselli).

https://doi.org/10.1016/j.najef.2025.102494

Available online 1 July 2025

National Academies of Sciences, Engineering, and Medicine, 2016), could affect financial markets and the economy as of today.

Natural gas has become one of the main energy sources in the world's leading economies (Hailemariam & Smyth, 2019). It is the least polluting hydrocarbon fuel and has lower carbon dioxide emissions compared to coal and oil. Additionally, it is increasingly being used as a precursor to hydrogen, which is seen as a potential alternative energy source for a low-carbon future. The U.S. is the world's largest NG producer and consumer (IEA, 2022), and the Henry Hub NG Futures is the second most actively traded energy futures contract based on a physical commodity worldwide.[1] However, the literature lacks a systematic analysis of the relationship between NG futures prices and weather information differentiating between observed weather and forecast weather information. This gap may be due to practical[2] and methodological[3] challenges in modeling NG futures, including the non-linear relationship with weather conditions and the probabilistic nature of weather forecasts. But given the increasing frequency of extreme weather events, understanding the link between NG future prices and weather, particularly extreme weather events, is crucial.

The central idea behind this study is that market participants who base their decisions on extreme weather forecasts face two sources of risk:

1. the first source of risk is tied to the time horizon of the forecast: weather forecasting error increases with the length of the forecast horizon;
2. the second source of risk is the "extreme" nature of the forecast itself: extreme weather events are rare and harder to predict compared to forecasts aligned with seasonal expectations.

As a result, acting on extreme weather forecasts over extended horizons entails a higher level of risk than forecasts aligned with seasonal norms. Financial markets are expected to efficiently reward this risk.

Our work builds upon this intuitive risk-reward concept by first addressing a crucial precursor: separately testing the effect of observed ("realized[4]") temperatures on NG futures compared to the effects of forecasted temperatures. Although not essential to the core analysis, this foundational testing is detailed in the Appendices to provide a comprehensive understanding of the impact of observed weather information as opposed to forecast weather information.[5]

From the premises of the informational efficiency results outlined in the Appendices, this study advances to explore the central premise: the existence of a risk premium in U.S. NG futures that is tied to extreme weather forecasts rather than to realized temperatures. This paper is structured as follows: Section 2 presents a review of the literature on NG future market and weather risk, the relationship between NG and crude oil and market efficiency in commodity markets; Section 3 guides the reader through the data and variables used in the analysis and their transformations; Section 4 discusses the results and findings. Finally, Sections 5 and 6 discuss and conclude the paper. Considering the variety of econometric tools employed and the range of hypotheses tested in this study, we present the following overarching summary to clarify for the reader how the identified risk premium is defined, identified, and captured, along with a high level description of the data used:

- the risk premium is defined as the "Abnormal Return" (AR) generated by extreme weather events within an Event Study framework;
- Extreme weather events are classified as those within the bottom 10 % of temperature forecast deviations from seasonal norms, a threshold shown in Section 4 to be flexible without loss of statistical significance;
- the risk premium is captured by a long and short position on the first and second NG futures contract, which also tend to be the most liquid NG contracts, when the following weather event condition applies within an Event Study framework[6]:

[1] The Henry Hub NG Futures is second only to the NYMEX West Texas Intermediate (WTI) Crude Oil futures: https://www.cmegroup.com/markets/energy/natural-gas/natural-gas.volume.html, https://www.cmegroup.com/markets/energy/crude-oil/light-sweet-crude.volume.html.

[2] Anecdotally, the NG industry calls the March-April NG futures spread the "widow maker". The Amaranth fund for example lost $5billion due to high volatility of the NG futures.

[3] As an example, several works either prove or refute NG future cointegration with Crude Light futures (see for example Ramberg and Parsons (2012) for a detailed analysis of the topic and the limitations caused by NG high volatility).

[4] Similarly to the difference in "realized volatility" as opposed to "implied volatility", we distinguish and test the effects of observed historical weather information as opposed to the effects of weather forecast on NG daily returns.

[5] Specifically:

- in Appendix A, we highlight the limitations of linear regression models in capturing weather effects, even when including typical NG market "fundamentals" such as storage levels and Gulf of Mexico natural gas extraction "shut-ins", as well as non-linear transformation of regressors. We then highlight with a quantile regression the non-linear response of NG 1-day returns to weather variables;
- in Appendix B we isolate the effects of daily and hourly "realized" temperatures on NG 1-day returns and separately tests the effects of forecasted temperatures using an Event Study methodology. This analysis confirms market efficiency in the classical sense, showing that observed temperature information does not produce statistically significant effects on NG daily returns while on the other hand forecast temperature have an impact on NG daily returns;
- Appendix C provides detailed tabulations of weather effects on NG futures daily returns from 1990 to 2019, with a control for False Discovery Rate (FDR) to ensure robust statistical validity.

[6] See Section 3 for weather variables definitions, Section 4 for a more complete definition of weather events and numerical results. Appendix B contains further analysis of additional configurations of weather events.

$$RiskPremimExtremeWeatherEvent_t \equiv temp_{diff_{t-1}}^{2Wforecast} < Quantile_{temp}^{10\%}$$

- the data used include high frequency observed temperature data (hourly frequency) and forecasted temperatures (daily frequency) across the major U.S. metropolitan areas spanning the 30-year period 1990–2019, sourced from the U.S. National Oceanic and Atmospheric Administration (NOAA).

  The key empirical findings are as follows:

- Observed temperatures have no significant daily impact on NG futures unless they deviate from the prior day's forecast.
- In contrast, forecasted sub-seasonal temperatures (particularly in autumn/winter) elicit strong market moves and drive a non-linear "weather-driven backwardation" on the futures curve.
- A simple futures spread strategy based purely on extreme weather forecasts effectively captures this "extreme weather" premium, outperforming the S&P 500 over 30 years, both in absolute returns and on a risk-adjusted basis.

The above findings have climate change and financial market implications as the rising frequency of extreme weather events may intensifies the relevance of the proposed "extreme weather" risk premium, hinting at potentially higher NG consumer costs and underscoring the importance of targeted risk management strategies in the energy sector focused on weather and climate risk.

2. Literature review

2.1. NG futures market and weather risk

NG futures prices are weather sensitive: the demand for NG is weather-driven and winter supply is constrained (Fleming et al., 2006). Roll (1984) was the first to suggest that the weather might influence financial markets, specifically futures markets: with reference to orange juice futures, he found a persistent impact of temperature shocks on price (i.e., a degree of market inefficiency). In general, the existing literature (e.g., Benth & Saltytė Benth, 2005, 2007; Bower & Bower, 1985; Cao et al., 2003; Considine, 2000; Elkhafif, 1996; Nick & Thoenes, 2014) identifies the volatility of temperature (i.e., temperature variations), including temperature shocks (Hu et al., 2014), as a factor affecting NG prices. Campbell and Diebold (2005) define the "weather risk" as the unpredictable component of weather fluctuations (i.e., weather surprises or weather noise), which needs to be considered in formulating hedging strategies (demand side) and in pricing derivatives (supply side). Mu (2007) incorporates these results with temperature shocks to assess their impact on NG price dynamics and demonstrates that temperature is a "more direct and purely exogenous measure of demand shocks".

Till and Di Tomasso (2000) and Till and Eagleeye (2005) refer to "weather fear premia" strategies and identify a type of risk premium: "A futures price will sometimes embed a fear premium due to upcoming, meaningful weather events". They note that there is positive expected value in systematically shorting certain futures contracts whose prices have built-in "weather fear premia" which later decrease if feared, but rare, weather events do not occur. The underlying logic of this strategy is that NG futures prices are viewed as systematically too high, reflecting the uncertainty of an upcoming weather event. Till (2008) notes that although over long periods of time it has been profitable to short weather-sensitive commodity markets at times of maximum weather uncertainty, these strategies can have very large one-off losses, which creates the classic "short option" risk/reward profile.[7] In relation to Till's "weather fear premia", the risk premium we identify is dissimilar (Till 2008, 2018). We show that there is actually a case for long short-term NG futures being based on weather forecasts, rather than being systematically short as proposed by Till.

Because demand is highly seasonal on the NG market (Todorova, 2004), inventories can be used to smooth production output and balance demand–supply conditions (Till, 2018). However NG storage is constrained by production output, which cannot adjust quickly to demand: this means that the information on NG storage is also important as a determinant of NG futures price volatility (Martinez & Torro, 2015; Shao et al., 2015). Any disruption in production, a hurricane for example, can boost NG prices where the storage level is constrained. Scott and Zhu (2004) analyze the short-term volatility of NG futures prices by observing the intraday prices for contracts traded on the NYMEX, and how price volatility is influenced by the American Gas Association's weekly Storage Survey, and the subsequent weekly report issued by the EIA (Energy Information Administration). Scott and Zhu (2004) find that publishing the reports increases volatility for up to 30 min. Chiou-Wei et al. (2014) show an inverse relation between the change in storage surprise (actual change minus expected change) and NG futures price changes on the day of the EIA storage announcement. Chiou-Wei et al. (2020) also suggest that market fundamentals, such as weather information, should be included in modelling commodity price movements, specifically for high weather sensitive commodities, such as NG. Finally, Liang et al. (2022) make the first successful attempt to directly incorporate extreme weather events into a GARCH framework: they show that their newly created GARCH-MIDAS-W-ES model including temperature and precipitation variables achieves the best accuracy in NG volatility forecasting. Liang et al. (2022) therefore suggest that it is necessary to analyze the role of extreme weather in volatility forecasting in the NG market. It must be stressed however that Liang et al. (2022) focus on NG1 volatility, while this paper focuses on NG futures daily returns. As shown by Andersen

[7] In other words, such strategies exhibit returns distributions with a positive skewness and "fat" left tails. They generate frequent small gains and infrequent but sometimes sizeable losses.

and Bollerslev (1998), volatility models produce strikingly accurate inter-daily forecasts, while daily financial returns are still to date to be predicted at a daily frequency.[8]

2.2. Relationship between NG and crude oil prices

In the energy industry, unofficial"rules of thumb" have long been used to link NG prices to crude oil prices. One of these was the "10-to-1 rule", i.e., NG price was one tenth of crude oil price. This rule evolved during the late 1990s and early 2000s to become the "6-to-1 rule"; in other words, the price of a million British thermal units of NG is roughly one-sixth of the price of a barrel of crude oil. These empirical "back of the envelope" calculations were superseded by the more complex "burner tip" rule, which takes into account price competition between petroleum products and NG where it actually occurs, at the burner tip. The burner-tip parity rule shows that NG prices generally track those of NYMEX WTI Crude Oil, but have decoupled from crude oil prices on many occasions. Many studies (e.g., Bachmeier & Griffin, 2006) show that NG and crude oil prices are cointegrated, and identify a long run equilibrium relationship. Hartley et al. (2008) examine the factors that cause short run departures from the long run equilibrium price relationship between NG and crude oil prices and find that the link between the two prices is indirect, and short run departures from long run equilibrium are influenced by product inventories, weather, other seasonal factors and supply shocks such as hurricanes. Ramberg and Parsons (2012) show that although the two-price series can be cointegrated, the confidence intervals for both short and long-time horizons are large. Brigida (2014) finds evidence for first-order Markov regime switching in the cointegrating relationship between NG and crude oil prices, meaning that the transition probabilities are the same over long time periods and they are not affected by economic and other conditions. Brigida (2016) investigates the determinants (i.e., storage, temperature, fuel demand and macroeconomic variables) of the endogenous regime switching process underlying the cointegrating relationship between NG and crude oil. Controlling for the effect of regime switching, Brigida (2016) observes that crude oil and NG prices are cointegrated, and that the cointegrating process is a function of factors affecting supply and demand for NG (e.g., storage). Brown and Yücel (2008) find that weekly oil and NG prices have a close relationship, but it is affected by a variety of exogenous and transitory factors such as seasonality, inventories, weather and production halts in the Gulf of Mexico (Breyer & MacAvoy, 1973). Other researchers on the other hand note that there is very low-price dependency between NG and crude oil (Batten et al., 2016) and even that there is no cointegration between the prices (Siliverstovs et al., 2005; Erdős, 2012; Lin & Li, 2015). In fact, there appears to be no consensus in literature on the relationship between NG and crude oil prices, and as noted by Chiou-Wei et al. (2020), weather, which is one the key exogenous variables impacting on NG, has been overlooked in many analyses to date.

2.3. Market efficiency in commodity markets and climate change considerations

Fama (1970), summarizing the empirical findings based on the Efficient Market Hypothesis (Fama, 1965; Samuelson, 1965a,b), states that efficient markets "always reflect all available information". On an efficient market, prices adjust instantaneously to any new piece of information so that all the available information is reflected in present prices (Fama, 1998).

Efficiency in futures markets is a fundamental concept, often defined by the zero expected net profit rule, which implies a fair and unbiased game (Bigman et al., 1983). According to the efficiency hypothesis, future prices should serve as unbiased estimates of spot prices at the delivery date, and there should be no consistent patterns in past forecast errors (Kellard et al., 1999). Examining the U.S. NG market, Gebre-Mariam (2011) specifically focuses on spot and futures contracts and finds that daily prices do not offer arbitrage opportunities, suggesting market efficiency. However, the study highlights that market efficiency is primarily observed for contracts with approximately one month until maturity, such as NG1 contracts.

The event study is a commonly recognized and established econometric tool used to assess market efficiency (Fama et al., 1969; MacKinlay, 1997, Khotari & Warner, 2006). It has been widely used to examine the impact of specific events on financial markets, allowing researchers to analyze abnormal returns and assess market efficiency. Several studies have utilized event study methodologies to examine market efficiency in the commodities market (Liu et al., 2019; Demirer & Kutan, 2010; Deaves & Krinsky, 1992). Liu et al. (2019) focus on OPEC announcements and their impact on crude oil price and volatility. Their findings reveal asymmetric price reactions to production decisions, with price increases having a more significant effect than price cuts. Kutan and Demirer (2010) investigate abnormal returns in crude oil spot and futures markets around OPEC conference and U.S. Department of Energy (DOE) Strategic Petroleum Reserve (SPR) announcement dates. They discover asymmetric price behavior, with statistically significant impacts observed only for OPEC production cut announcements, and the impact diminishing for longer maturities. Deaves and Krinsky (1992) observe significant excess returns for market participants who went long in oil futures contracts following the conclusion of OPEC conferences; they also note that these excess returns exceeded those typically found in equity market event studies.

To the best of our knowledge, there is an absence of event studies examining the impact of weather events on financial markets like the one proposed here. Furthermore, as noted in Section 1.1, there is in the existing literature a lack of studies investigating the

[8] Malkiel and Fama (1970) examine various forecasting models and find that short-term return predictability is generally weak or non-existent. Consistent with this, Fama and French (1988) conclude that stock returns show limited predictability beyond what can be explained by market risk factors. Campbell and Yogo (2006) expand the analysis by exploring a wide range of predictors and find that while some exhibited statistical significance, their economic value in predicting daily returns is limited. Empirical research has shown that predicting daily commodity returns is a challenging task, similar to stock returns: Gorton and Rouwenhorst (2006) find limited evidence of predictable patterns in commodity futures returns, with economic risk factors accounting for much of the performance.

efficiency of NG futures market movements in response to weather information that differentiate between observed and historical weather information as opposed to forecast weather information. This study aims to fill this gap and shed light on the relationship between weather events and the efficiency of the NG futures market.

The topic addressed in this study is important in the light of the growing body of evidence pointing to the impact of climate change on weather patterns. Numerous studies (Cohen et al., 2018; Cohen et al., 2021; Francis et al., 2012; National Academies of Sciences, Engineering, and Medicine, 2016) have highlighted the increasing occurrence of extreme weather events as a consequence of climate change. Understanding the reaction of NG futures to weather fluctuations, especially extreme fluctuations, has become essential. This research aims to contribute to the existing literature by investigating the relationship between weather disruptions driven by climate change and the behavior of NG futures. By examining the market's response to extreme weather events, this study provides valuable insights into the implications of climate change for the NG futures market and the broader energy sector.

3. NG and weather data

In this study, we construct an *alternative* dataset of U.S. temperatures and link it with traditional economic data. The aim of this alternative dataset is to transform multidimensional weather data into variables manageable with classical econometric time-series analysis tools. Table 1 provides a summary of the datasets used in this study.

3.1. Data and variables definitions

3.1.1. Market data

Market data for NG prices and NG "market fundamentals" (i.e., the variables identified in literature as the main drivers of NG futures prices) considered in our analysis are the following:

i. Daily prices set in USD of NG futures contracts with expiry: 1 month ("near-month", "front-month"), 2 months ("second expiry") (Source: NYMEX, retrieved through Bloomberg);
ii. Daily prices set in USD of crude oil (WTI) futures contracts with expiry1 month ("near-month", "front-month"), 2 months ("second expiry") (Source: NYMEX, retrieved through Bloomberg);
iii. Weekly U.S. NG Storage[9] in Billion Cubic Feet (BCF) as reported by the U.S. EIA (retrieved through Bloomberg);
iv. Gulf of Mexico NG production shut-downs (Source: Mineral Management Service, U.S. Department of Interior).

3.1.2. Weather data

According to Auffhammer et al. (2013), weather and climate differ in terms of time: weather refers to atmospheric conditions over a short period of time, and climate reflects the behavior of the atmosphere over a relatively long period of time.[10] The aim of the construction of a dedicated dataset was to make it possible to aggregate "spatially dispersed and correlated weather data" (Auffhammer et al., 2013) into single proxies for the entire U.S. with the purpose of linking these weather variables to financial market data.

We pieced together the following U.S. weather datasets (see Table 1 for a recap):

i. **Hourly U.S. Observed ("realized") Temperatures:** hourly U.S. historical temperature[11] readings from the 12 biggest U.S. metropolitan areas,[12] measured at the reference airport weather stations (see Table 2) from 1 August 1989 to 31 July 2019. We obtained this dataset from the NOAA's National Centers for Environmental Information (NCEI), formerly the National Climatic Data Center – NCDC. These data are used to build U.S. temperature time series at 14:00 EST, 30 min before the daily settlement time for NG futures contracts (14:30 EST), which is used to assess the effect of actual observed temperatures levels across the U. S. on NG futures;
ii. **Daily U.S. Forecasted Temperatures:** daily temperature[13] forecasts for the 12 biggest U.S. metropolitan areas (see Table 2 and Fig. 1), from 1 January 1987 to 31 July 2019.[14] This dataset is sourced from the U.S. NOAA Global Ensemble Forecasting System – GFES (see Hamill et al. 2013). These data are used to build three U.S. temperature forecast time series, with forecast horizons of 1 day, 1 week and 2 weeks, and to assess the effect of temperature forecasts on NG futures;

[9] EIA "Lower 48 States Natural Gas Working Underground Storage".
[10] In their analysis of the use of weather data for economic analyses of climate change, Auffhammer et al. (2013) outline the major pitfalls of using weather data products in econometric models.
[11] Air temperature at two meters from the ground (dry-bulb temperature).
[12] The 12 biggest metropolitan statistical areas in the U.S., in decreasing order of total population as per U.S. Census 2010.
[13] Air temperature at two meters from the ground (dry-bulb temperature).
[14] The NOAA Physical Sciences Laboratory 2nd-generation Reforecast Project has produced a dataset of historical weather forecasts generated with a fixed numerical model, using the 2012 version of NCEP's Global Ensemble Forecasting System (GEFS, Version 10). The forecast data used in this study corresponds to the Reforecast V2 dataset, which consists of an 11-member ensemble of forecasts, produced every day from 00:00 UTC initial conditions from Dec 1984 to the present. The horizontal resolution of GEFS is T254 (about 50 km) for forecasts up to 8 days, and T190 (about 70 km) for forecasts between 8–16 days.

Table 1
Summary of the datasets.

Dataset	Variables description	Source	Period	Frequency	Use
Hourly U.S. Observed Temperatures	Hourly U.S. historical temperature readings dry-bulb temperature from the 12 biggest U.S. metropolitan areas	NOAA's National Centers for Environmental Information (NCEI), formerly the NCDC	1989–2019	Hourly	Testing NG NYMEX market efficiency towards observed temperatures.
Daily U.S. Observed Temperatures	Daily U.S. historical temperature readings from the 12 biggest U.S. metropolitan areas, measured at the reference airport weather stations	NOAA's NCEI, formerly the NCDC	1950–2019	Daily	Build climatological expectations for the period 1990–2019
Daily U.S. Forecasted Temperatures	Daily temperature forecasts for the 12 biggest U.S. metropolitan areas at 1 day, 1 week and 2 weeks forecast horizon.	NOAA Global Ensemble Forecasting System –GFES – (see Hamill et al. 2013)	1987–2019	Daily	Test the presence of an Extreme Weather Forecasts risk factor
NG: markets and fundamentals	Daily Henry HUB Natural Gas futures settlement price	NYMEX	1990–2019	Daily	Test the presence of an Extreme Weather Forecasts risk factor
	Daily Crude oil (WTI) futures settlement prices	NYMEX	1990–2019	Daily	Econometric model for NG returns
	Weekly U.S. NG Storage in Billion Cubic Feet (BCF)	U.S. Energy Information Administration (EIA)	1990–2019	Weekly	Econometric model for NG returns
	Gulf of Mexico NG production shut-downs in BCF	Mineral Management Service, U.S. Department of Interior	1990–2019	Ad-hoc	Econometric model for NG returns
	Daily S&P500 index price	Bloomberg	1990–2019	Daily	Control variable.

Table 2
Temperature measurement station location in the 12 biggest U.S. metropolitan areas.

Metropolitan statistical area	Weather measurement location	NCDC weather station ID	Latitude	Longitude
New York – Newark – Jersey City	JFK International Airport	WBAN: 94,789	40.64	−73.78
Los Angeles – Long Beach – Anaheim	Los Angeles	WBAN: 23,174	34.05	−118.25
Chicago – Naperville – Elgin	Chicago O'Hare International Airport	WBAN: 94,846	41.79	−87.75
Dallas – Fort Worth – Arlington	Dallas FAA International Airport	WBAN: 03,927	32.85	−96.85
Philadelphia – Camdem – Wilmington	Philadelphia International Airport	WBAN: 13,739	39.87	−75.24
Houston – The Woodland – Sugar Land	Houston William Hobby Airport	WBAN: 12,918	29.65	−95.28
Washington – Arlington – Alexandria	Washington Dulles International Airport	WBAN: 93,738	38.95	−77.46
Miami – Fort Lauderdale – Pompano Beach	Miami	WBAN: 12,839	25.80	−80.29
Atlanta – Sandy Springs – Alpharetta	Atlanta Hartsfield International Airport	WBAN: 13,874	33.78	−84.52
Boston – Cambridge – Newton	Boston	WBAN: 14,739	42.37	−71.01
San Francisco – Oakland – Berkeley	San Francisco	WBAN: 23,234	37.62	−122.38
Detroit – Warren – Dearborn	Detroit Metro Airport	WBAN: 94,847	42.41	−83.00

This table reports the location of temperature forecast stations for the hourly and daily temperature measurements provided by the NOAA's NCEI (Section 3). Regarding the daily temperature forecasts provided by the NOAA Global Ensemble Forecasting System – GFES – (Hamill et al. 2013), locations shown in this table have an approximation spatial error of up to 43 miles compared to the Global Ensemble Forecasting System forecast points. The temperature measurements and forecasts at the locations listed above were used to obtain weighted average temperature time series to proxy the U.S. temperature over the period 1990–2019.

iii. **Daily U.S. Observed ("realized") Temperatures**: daily U.S. historical temperature[15] readings from the 12 biggest U.S. metropolitan areas, measured at the reference airport weather stations (see Table 2), from 1 January 1950 to 31 July 2019. We sourced this dataset from the NOAA's NCEI. This data is used to build the U.S. long-term temperature seasonal averages for each day of the year: these day-of-the-year long term average temperatures are also known as *climatological expectations*, *seasonal expectations,* or *seasonal norms*.

3.1.3. *Construction of econometric weather variables*

Our analysis aimed to condense the complex weather information into single variables by making two strong assumptions:

1) we focused on temperature as it is widely acknowledged as one of the strongest climatological explanatory variables because of its key role in shaping the Earth's physical and biological systems. Temperature is one of the most important weather variable for economic activity in the U.S. as it drives directly the heating demand. Heating demand drives mainly fossil fuels demand, including

[15] Air temperature at two meters from the ground (dry-bulb temperature).

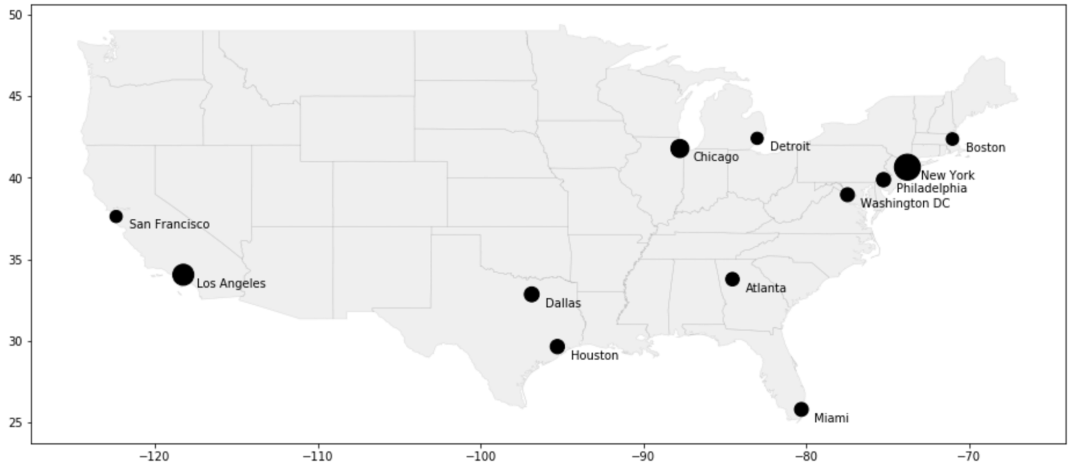

Fig. 1. Twelve temperature measurement locations across the U.S. Fig. 1 depicts the geographical distribution of temperature measurement stations utilized in this study to derive observed and forecasted average U.S. temperature time series. The size of each dot reflects the relative population size of the corresponding metropolitan statistical area, as reported in the 2010 U.S. Census. This population weighting approach enabled us to construct spatially weighted average time series of temperatures, representing the 12 largest metropolitan areas in the U.S. These weighted average time series were then utilized to proxy the daily temperature dynamics across the entire U.S. during the period from 1990 to 2019.

natural gas. When dealing with weather and climatological variables in relation to short and long term economic activity, researchers often focus on temperature given its importance as weather variable such as Dell et al. (2012), Kotz et al. (2021). When analyzing Natural Gas specifically, many researchers have focused on temperature like Mu (2007), Song et al. (2015), and utilize temperature as the main weather variable in their econometric models.

2) we spatially averaged temperature time series across the U.S., using the population size of the 12 U.S. largest metropolitan statistical areas as a basis for weighting.

This approach allowed us to capture the temperature conditions affecting the majority of residential and industrial areas in the U.S. The 12 largest metropolitan areas in the U.S. account for approximately 37 % of U.S. GDP in 2019.[16] Fig. 1 provides a graphical representation of the metropolitan areas and ponderation weights[17] for the weather time series described in this section.

Using this methodology, we obtained the following U.S. daily temperature proxy time series for each business day in the period 1990–2019:

i. $\boldsymbol{temp_t^{ACTUAL14:00}}$: U.S. population weighted average temperature as observed at 14:00 EST of any business day *t*, in degrees Celsius (°C). For each historical date (denoted as *t*), the variable can be interpreted as the U.S. temperature observed 30 min before the determination of the NYMEX settlement price of the NG futures, which occurs at 14:30 EST. This variable is shown as the dotted line in Fig. 2;

ii. $\boldsymbol{temp_t^{1dforecast}}$: U.S. population weighted average 1-day ahead forecasted temperature in °C. The one day ahead forecasted temperatures of each of the 12 biggest U.S. metropolitan areas for any business day *t* are sourced from the NOAA's GFES based on 00:00 UTC[18] initial conditions at business day (*t*). For each historical date *t*, the variable can be interpreted as the average U. S. point in time temperature forecast at *the end of the next business day (t + 1 business days)*;

iii. $\boldsymbol{temp_t^{1Wforecast}}$: U.S. population weighted average 1-week ahead forecasted temperature in °C. The one week ahead forecasted temperatures of each of the 12 biggest U.S. metropolitan areas for any business day (t) are sourced from NOAA's GFES based on 00:00 UTC initial conditions at business day *t*. For each historical date *t*, the variable can be interpreted as the average U.S. point in time temperature forecast at *the end of the next week (t + 5 business days)*;

iv. $\boldsymbol{temp_t^{2Wforecast}}$: U.S. population weighted average two week ahead forecasted temperature in °C. The two week ahead forecasted temperatures of each of the 12 biggest U.S. metropolitan areas for any business day *t* are sourced from NOAA's GFES, based on 00:00 UTC initial conditions at business day *t*. For each historical date *t*, the variable can be interpreted as the average U.S. point in time temperature forecast at *the end of the next 2 weeks (t + 10 business days)*.

v. $\boldsymbol{temp_t^{seasonalexpectation14:00}}$: referred to as the U.S. population weighted average seasonal norm, represents the long-term temperature average for each day of the calendar year associated with day t, in °C. The long-term temperature averages are first computed for each of the 12 U.S. metropolitan areas, using a 50-year average of the temperature for every day of the year. The

[16] https://www.bea.gov/news/2020/gross-domestic-product-county-2019.

[17] The averaging was obtained with ponderation weights corresponding to the U.S. metropolitan areas population as per 2010 Census.

[18] Note that Eastern Standard Time (EST) is 5 h behind Coordinated Universal Time (UTC).

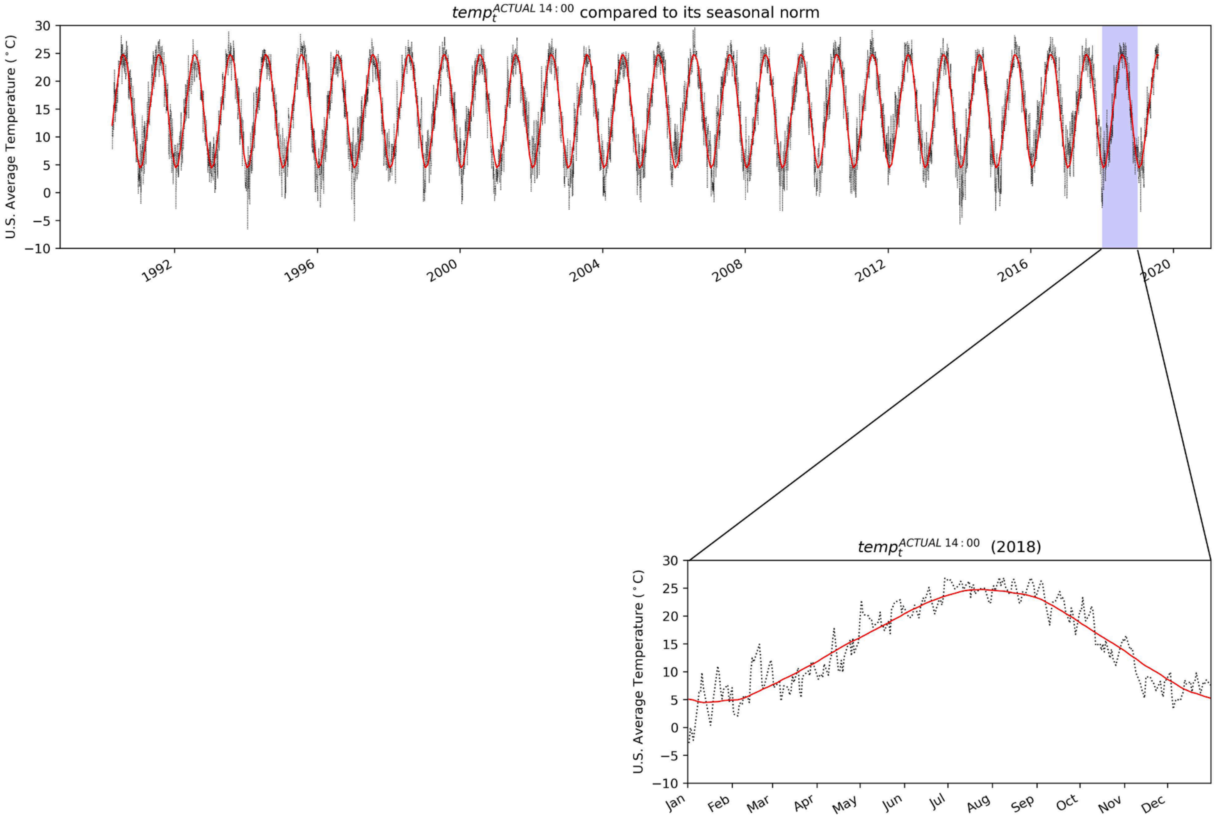

Fig. 2. U.S. daily average temperature observed at 14:00 EST compared to seasonal norms with enlargement showing details for 2018. This graph shows U.S. daily temperatures (weighted averages of the 12 biggest metropolitan areas) measured at 14:00 EST ($temp_t^{ACTUAL14:00}$, dotted line) compared to climatological seasonal expectation ($temp_t^{\text{seasonal expectation }14:00}$, red line), enlarged for year 2018. Temperatures are expressed in degrees Celsius. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

population-weighted average of these averages is then taken to obtain a U.S. population-weighted climatological expectation for each day of any calendar year at 14:00 EST. The variable can be interpreted as the average U.S. climatological expectation or seasonal norm for each day of the year. This variable is the continuous red line in Fig. 2, showing the difference between short-term "weather" fluctuations and long-term climatological expectations.

3.1.4. Normalization of weather variables against seasonal expectations

In the field of climatology, seasonal averages, also known as *climatological expectations*, *seasonal expectations*, or *seasonal norms*, are widely considered the most reliable long-term forecasts (Enger, 1959). These seasonal climatological expectations should also be factored into market prices, as suggested by Borovkova and Geman (2006).

In the context of weather-sensitive markets such as NG futures, forecasts that significantly differ from already anticipated weather conditions should have a market-moving impact. To capture the impact of short-term weather fluctuations on markets, we therefore want to normalize our weather time-series by subtracting the seasonal expectation from the original weather time series. The objective is to capture unplanned energy demand and ascertain the effect of short-term weather fluctuation on markets, similarly to what is done with Heating degree day (HDD) and Cooling degree day (CDD) indices. This approach allows us to consider our weather variables in terms of differentials from seasonal expectations. Our variables normalized for long term climatological expectations are computed as follows:

vi. $\boldsymbol{temp\_diff}_t^{\boldsymbol{actual}14:00} = \boldsymbol{temp}_t^{\boldsymbol{ACTUAL}14:00} - \boldsymbol{temp}_t^{\textbf{seasonal expectation } 14:00}$: the difference between the U.S. temperature at 14:00 EST and its climatological expectation;

vii. $\boldsymbol{temp\_diff}_t^{1\boldsymbol{Dforecast}} = \boldsymbol{temp}_t^{1\boldsymbol{Dforecast}} - \boldsymbol{temp}_{t+1}^{\textbf{seasonal expectation}}$: U.S. one week ahead forecasted temperature difference from climatological expectations;

viii. $\boldsymbol{temp\_diff}_t^{1\boldsymbol{Wforecast}} = \boldsymbol{temp}_t^{1\boldsymbol{Wforecast}} - \boldsymbol{temp}_{t+5}^{\textbf{seasonal expectation}}$: U.S. one week ahead forecasted temperature difference from climatological expectations[19];

ix. $\boldsymbol{temp\_diff}_t^{2\boldsymbol{Wforecast}} = \boldsymbol{temp}_t^{2\boldsymbol{Wforecast}} - \boldsymbol{temp}_{t+10}^{\textbf{seasonal expectation}}$: U.S. two weeks ahead forecasted temperature difference from seasonal expectations.[20]

The variable $temp\_diff_t^{actual14:00}$ will be used as proxy for the "*U.S. observed temperature at 14:00 EST differences from long term seasonal climatological expectations*".

The variables $temp\_diff_t^{1Dforecast}$, $temp\_diff_t^{1Wforecast}$ and $temp\_diff_t^{2Wforecast}$ will serve as proxy for "*U.S. forecasted short-term weather differences from long term seasonal climatological expectations*" with forecasting horizons 1 day, and 1 and 2 weeks, respectively.

Fig. 3 displays the U.S. observed temperature at 14:00 EST time series (shown previously in Fig. 2), after the subtraction of the seasonal climatological norms as described in points vi-ix above.[21]

A summary of native and normalized weather variables used in this study is provided in Table 3. Table 4 shows the descriptive statistics of the U.S. daily temperatures normalized for seasonal expectations described above, using the daily data from 1990 to 2019. It is important to note that the temperature differentials from seasonal norms appear to be mean and variance stationary, as confirmed by stationarity testing (see Appendix A). Note also that when the normalized temperature variable equals zero degree Celsius, this indicates that the current or forecasted temperature is in line with its seasonal expectation.

3.2. Short term weather forecasts as stochastic variables

It is widely accepted that the best long-term forecasts in climatology are seasonal averages (Enger, 1959) and that therefore long-term seasonal climatological norms are incorporated into market prices (Borovkova & Geman, 2006). However, short-term weather fluctuations around these norms cannot be widely incorporated well in advance. The future realization of temperature above or below its seasonal expectation can be seen as the realization of a stochastic variable: temperature forecasts are inherently probabilistic (Lorenz, 1963) and the NOAA forecast variables we are using are ensemble probabilistic forecasts (Hamill et al., 2013). Fig. 4 shows the correlation between our single-variable U.S. temperature forecast and ex-post temperature observations made across the U.S. from 1990 to 2019. Although the forecasts appear to be reliable during the period 1990–2019, there is still a significant forecasting error component.

In analyzing the relationship between temperature forecasts and actual future ex-post temperatures, we observe that the error distributions tend to be mostly symmetrical with increasing error magnitude, as shown in Fig. 5. It is trivial but important to note that the risk associated with forecasting error increases as the forecast horizon extends further into the future, especially when such

[19] Only business days are considered, so the 1 week ahead forecasted temperature is matched with the "t+5 business days" seasonal long-term average, which corresponds to "t+7 calendar days".

[20] Only business days are considered, so the 2 weeks ahead forecasted temperature is matched with the "t+10 business days" seasonal long-term average, which corresponds to "t+14 calendar days".

[21] This normalization of weather variables is very similar to the methodology underlying the Heating degree day (HDD) and Cooling degree day (CDD) indices.

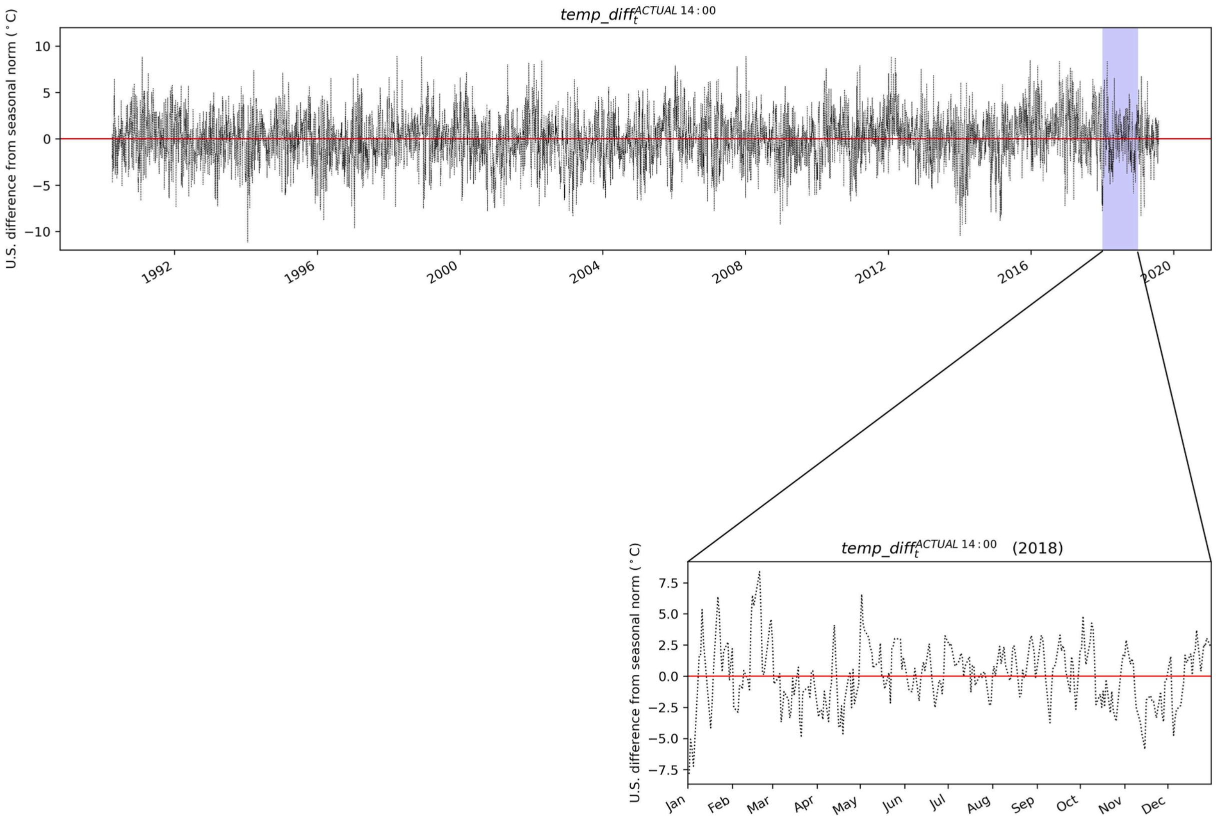

Fig. 3. U.S. daily average temperature observed at 14:00 EST after normalized for seasonal expectations. This graph shows of the U.S. daily temperatures (weighted averages of the 12 biggest metropolitan areas) measured at 14:00 EST ($temp_t^{ACTUAL 14:00}$, dotted line) once normalized to its seasonal expectations ($temp_t^{\text{seasonal expectation 14:00}}$, red line) with the enlargement showing 2018 in more detail. Temperatures are expressed in degrees Celsius. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of native and normalized weather variables used in this study.

Variable	Description
$temp_t^{seasonalexpectation14:00}$	Climatological expectation for the U.S. temperature at 14:00 EST
$temp_t^{ACTUAL14:00}$	Temperature at 14:00 EST
$temp_t^{1dforecast}$	One day ahead forecasted temperature
$temp_t^{1Wforecast}$	One week ahead forecasted temperature
$temp_t^{2Wforecast}$	Two weeks ahead forecasted temperature
$temp\_diff_t^{actual14:00}$	Observed temperature at 14:00 EST difference from its climatological expectation
$temp\_diff_t^{1dforecast}$	One day ahead forecasted temperature difference from its climatological expectation
$temp\_diff_t^{1Wforecast}$	One week ahead forecasted temperature difference from its climatological expectation
$temp\_diff_t^{2Wforecast}$	Two weeks ahead forecasted temperature difference from its climatological expectation

This table shows all weather variable used in this study, referring to the 12 U.S. biggest metropolitan areas weighted average. See Section 3.1.4 for further details on the weather variables adjusted for seasonal expectations.

forecasts are used to inform economic decision-making.

3.3. Short term weather forecasts and NG price correlation

We observe significant level correlation between NG1 prices and weather variables (Fig. 6). This correlation is however time varying, pointing to potential non-linearities. In Appendix A we indeed show with quantile regressions that NG returns show a highly non-linear relation with forecasted temperatures. We also analyze the potential relationship between NG1 returns and weather variables through regression analysis: the analysis shows that NG1 returns are linearly dependent on 1-week ahead U.S. forecasted temperatures, but they do not appear to be dependent on U.S. realized temperatures levels (at 14:00 EST).[22] Although statistically significant, the regression shows low predictive power with an R2 of 4.7 %, which is remarkably similar to Roll's R2 from his findings on the explanatory power of weather on the orange juice market (Roll, 1984).

3.4. Data reassurance: data relevance, market liquidity, avoidance of data leakages and study replication considerations

Since we use several "alternative" non-standard datasets (i.e. U.S. observed and forecasted temperatures), the risk of data revisions, forward-looking biases in the historical estimates and issues in term of relevance of the choice of variables may raise (e.g., NG first and second expiry futures). Below we summarize data reassurances points and methodological choices to rule out different data issues and to support the replication of this study:

- Data relevance: we use weather variables as weighted averages of the 12 biggest U.S. metropolitan areas using as weights the population from 2010 U.S. census. To use a single snapshot of the U.S. population as of 2010 introduces forward-looking bias in the weights used: a replication of the study was performed with equal weights to rule out this risk;
- Market liquidity: we chose to work on NG futures first (NG1) and second nearby expiries (NG2) as those futures are usually the most liquid expiries for commodity futures. Although longer dated NG futures have relevant volumes and open interests, usually the first and second expiry have the most of market liquidity[23]: this is the same approach followed by Boons and Porras Prado (2019).[24]
- Data leakages (forward-looking bias): this is a risk associated with any back-tested market strategy where the signal variable (weather forecast in our case) is temporally overlapped with the return computation window of the traded asset. We ruled out this error as follows:
  - o For the test of the market efficiency of NG futures against observed temperatures described in Appendix B, we used hourly observations to derive a daily temperature variation specifically aligned to the NG futures daily returns computation time window (see Fig. B.1);
  - o For the computation of the performance of the risk factors we highlight that NOAA weather forecasts in our dataset are released at 00:00 Coordinated Universal Time (UTC). The Nymex exchange where the NG futures are traded is in New York (Eastern Standard Time (EST) time zone) which is five hours behind Coordinated Universal Time (UTC − 05:00): this means that the forecast we use are released as public information at 19:00 EST. Natural gas futures contracts are now traded on the NYMEX on a

[22] The only temperature related coefficient which is statistically significant is that associated with the forecasted temperature, while we found that the 14:00 EST realized temperature appears to not influence NG1 returns in the regression analysis. This was tested using the whole dataset and also while running the OLS regression on different seasonal subsets of the dataset, i.e., only considering winter months.

[23] Live volume and Open Interests for NG futures are published daily on the CME website: https://www.cmegroup.com/markets/energy/natural-gas/natural-gas.volume.html.

[24] "We define basis-momentum as the difference between momentum in a first- and second-nearby futures strategy", Boons and Prado (2019).

Table 4
Descriptive statistics of the U.S. daily temperature differentials from seasonal norms.

Variable	$temp\_diff_t^{actual14:00}$	$temp\_diff_t^{1Dforecast}$	$temp\_diff_t^{1Wforecast}$	$temp\_diff_t^{2Wforecast}$
Description	Actual U.S. temperature at 14:00 EST difference from seasonal expectation	U.S. one day ahead forecasted temperature difference from seasonal expectation	U.S. 1 week ahead forecasted temperature difference from seasonal expectation	U.S. 2 weeks ahead forecasted temperature difference from seasonal expectation
Mean	0.0	0.0	0.0	0.0
Std	2.6	2.5	2.6	3.2
Min	−11.1	−11.6	−9.7	−10.3
25 %	−1.6	−1.4	−1.6	−2.7
50 %	0.0	0.2	0.4	0.6
75 %	1.7	1.7	1.9	2.9
Max	9.0	8.9	7.7	6.7
Skew	−0.1	−0.4	−0.5	−0.3
Kurtosis	0.5	0.6	0.1	−1.0

This table shows descriptive statistics of the daily time series of U.S. weighted average daily temperatures difference from seasonal expectations, in degrees Celsius (daily data, 1990–2019). N. observations: 7,423.

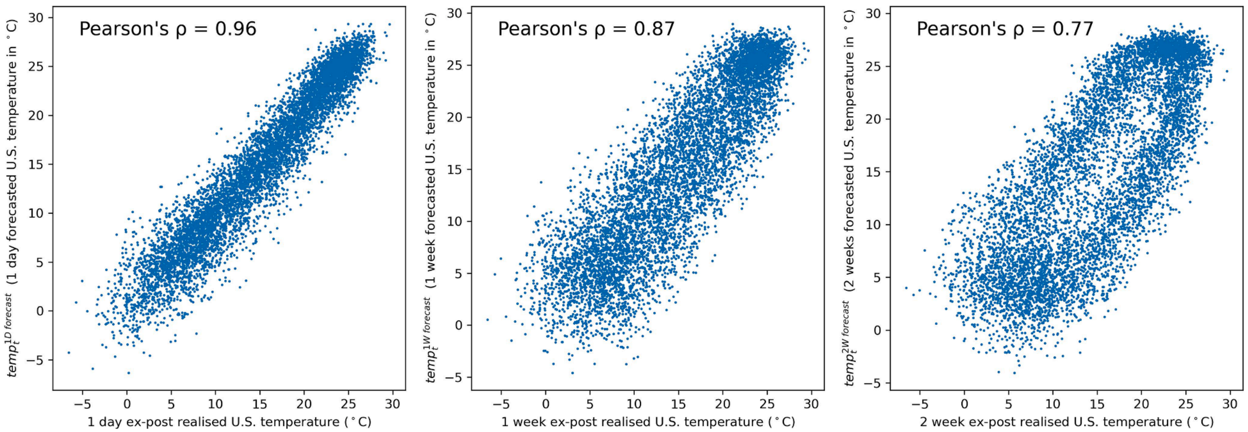

Fig. 4. U.S. daily average temperature forecasts vs ex-post realized temperatures, 1990–2019 In the Figure, each dot corresponds to the U.S. temperature (weighted average of the 12 biggest U.S. metropolitan areas forecasted and ex-post realized temperature. The figure shows that NOAA publicly available forecasts (y axis) tend to be quite reliable compared with the ex-post temperature readings (x axis), even after aggregating the forecast from several U.S. locations with different local climatological conditions as we did in this work.

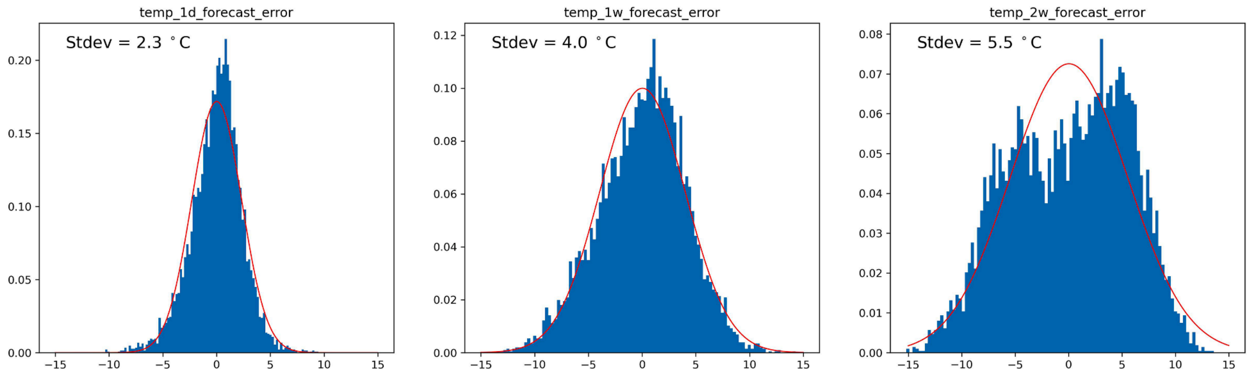

Fig. 5. U.S. daily average temperature forecast error distribution, 1990–2019. The graph shows how the magnitude of the ex-post forecast error increases with the forecast horizon, with Gaussian fit overlay for comparison purposes. The distributions of the ex-post forecasting error consider single weather forecasts time series for the whole U.S. (see Sections 3.1.4), covering the period 1990 to 2019 using NOAA forecasting data.

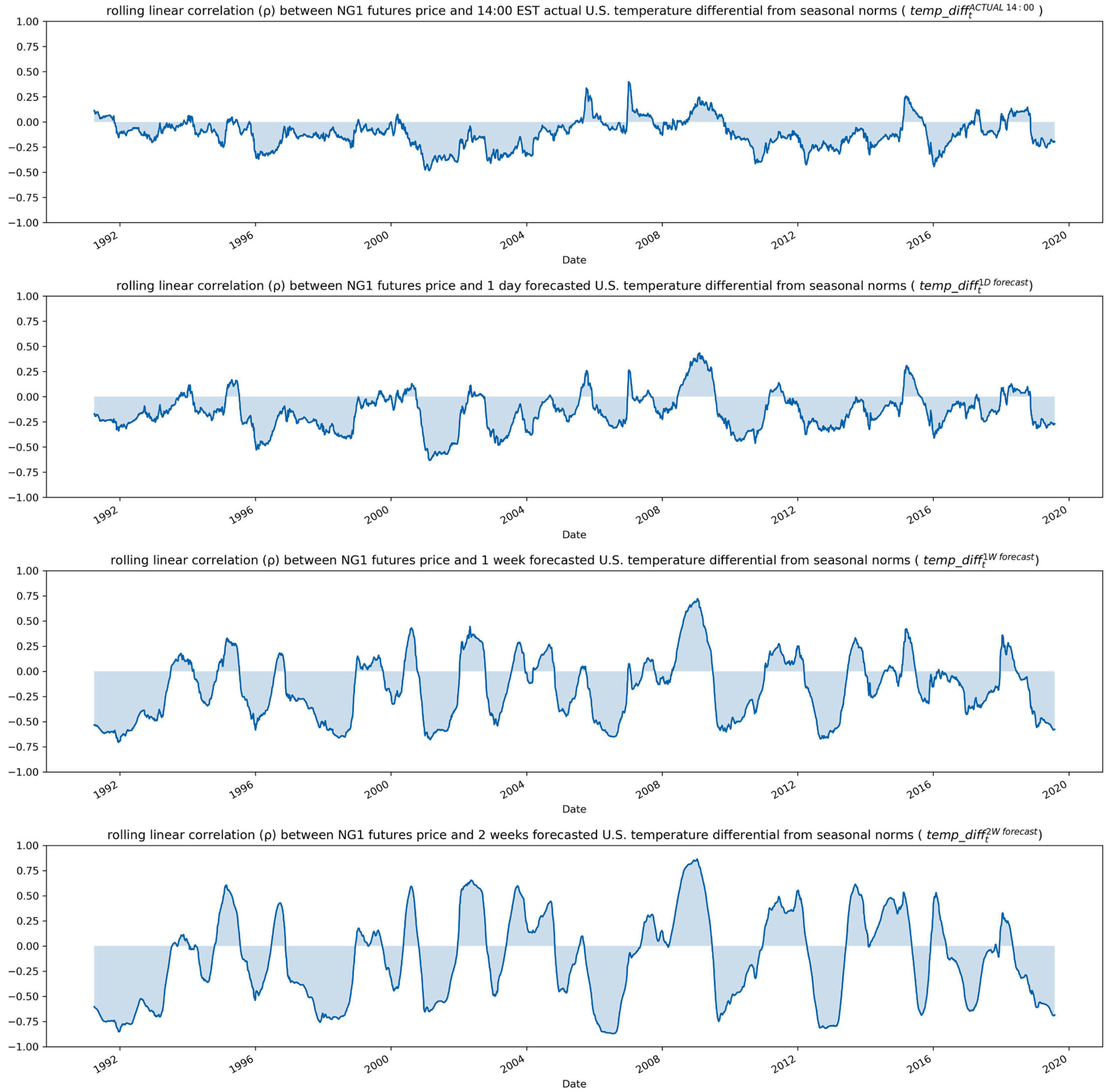

Fig. 6. Rolling 1-year correlation between NG1 prices and U.S. daily average temperatures. Rolling 1 year (260 days) correlation between NG futures front-contract (NG1) prices and U.S. temperature differentials from seasonal expectations: observed 14:00 EST temperature $temp\_diff_t^{\mathrm{ACTUAL}\ 14:00}$, forecasted 1-day ahead $temp\_diff_t^{1D\ forecast}$, forecast 1-week ahead $temp\_diff_t^{1W\ forecast}$, forecast 2-weeks ahead $temp\_diff_t^{2W\ forecast}$. The correlation increases with forecast horizon and shows a time-varying dynamic. Based on daily data from April 4, 1990 to July 31, 2019.

24 h basis, however in order to be consistent with historical quotation standards, we computed returns from settlement price to settlement price, which is set at 14:30 EST. To compute the spreading strategy returns described in the next section, we use $t + 1$ returns, meaning that we leave a 19 ½ hours of time between the weather forecast signal and the start of the computation of the return of the trades that compose the spreading strategy.

In order to facilitate the study replication, we provide upon request both the dataset and a Jupyter™ Notebook in Python™ language with a full replication of the study.

4. Empirical analysis: a novel "Extreme Weather Forecast" risk premium

Our findings in Appendix B reveal that the impact of observed ("realized") temperatures on NG daily returns is negligible except for one particular case,[25] while forecasted temperatures do have an impact on NG returns. In the U.S., weather forecasts have been in the public domain for several decades and are readily available to the market from government agencies such as the NOAA and the NWS. The availability of weather forecasts has allowed market participants to incorporate weather-related information into their trading decisions, particularly in sectors such as energy, agriculture, and transportation, where weather can have a significant impact on prices and demand (Easton et al., 2017, Huurman et al., 2012., McGuirk et al., 2009, Turvey, 2001).

4.1. *"Extreme Weather Forecast" risk premium: proposed economic rationale and hypothesis tested*

Chang (1985) defines "risk premium" as "[…] an average reward to investors for being willing to assume a risk position in a risk-averse financial world. The reward in this form should not be conditioned on any superior judgment or inside information."

A proposed economic mechanism linking extreme weather events and NG prices proposed here works as follows: when temperatures drop (rise) significantly below (above) seasonal norms, energy consumption increases as consumers turn up their thermostats (air conditioning) to maintain comfortable indoor temperatures. This leads to an unplanned increase of short-term energy demand, which leads to an increase in demand for NG. This increased demand for NG leads to a price premium for the nearest NG future delivery ("front-end" contract, NG1) as market participants anticipate higher demand for the physical NG due to the increased demand for energy. This mechanism however is based on the assumption that market participants efficiently and precisely price weather risk for NG but given the probabilistic nature of weather forecasting this is possibly not the case. Weather forecasts are in fact intended to reduce the uncertainty stemming from short-term weather fluctuations, but they also bear intrinsic model risk (i.e. forecasting error). There are two sources of intrinsic risk: NG market operators face "model risk" from the accuracy level of weather forecasts as well as "physical world risk" from weather fluctuations. Financial markets always reward for risk[26] and our findings suggest that this combination of physical world risk and model risk condenses into an "extreme weather forecast" risk premium.

Temperature forecasts are probabilistic estimates subject to intrinsic model errors (Section 2). The error associated with temperature forecasting tends to increase as the length of the forecasting horizon increases (Fig. 5) up to a point in time in the future where the forecasts completely lose any predictive power due to the chaotic nature of atmospheric flows (Lorenz, 1963). Given these intrinsic limitations, it is logical to expect that economic agents taking market positions based on weather forecasts are always subject to a degree of weather forecasting error. This means that the accuracy of weather forecasts will always be a factor that market participants must take into account in investment decisions. It also means that model risk cannot be completely eliminated when making investment decisions based on weather forecasts. It appears therefore logical that this ineliminable risk should be rewarded by a risk premium, which should compensate market participants for the uncertainty and potential losses associated with incorrect predictions of weather-related events.

We have already noted that the forecasting error increases as the forecasting horizon increases, especially during winter when temperature volatility is higher (Fig. 5). However, a less known fact is that extreme temperature forecasts tend to have a higher forecasting error compared to forecasts that are more aligned to seasonal norms. Fig. 7 shows that the ex-post forecasting error increases as the level of forecasted temperatures becomes more extreme (i.e. not aligned to seasonal norms, which corresponds to the zero point on the x-axis in Fig. 7). This effect is amplified as the forecasting horizon lengthens and suggests that as the time frame for forecasting time horizon lengthens, the uncertainty surrounding more extreme temperature forecasts becomes more pronounced. Consequently, market participants who base their decisions on extreme weather forecasts for a longer horizon face two sources of uncertainty: he first one is tied to the time horizon of the forecast; the second one is the "extreme" nature of the forecast itself.

As a result, acting upon extreme weather forecasts entails a higher level of risk compared to forecasts that align with seasonal expectations (represented by x-axis and y-axis origin in Fig. 7).

Our proposed economic rationale regarding the behavior of a weather forecast-related risk premium can be summarized as follows:

i. longer forecasting horizons are associated with higher forecasting errors;
ii. for a given forecasting horizon, more extreme forecasts exhibit greater forecasting errors;
iii. when considering both a longer forecasting horizon and a more extreme forecast, the level of error is expected to be higher.

On these premises, we predict that the performance of any risk premium linked to weather forecasts will exhibit a monotonically increasing magnitude in response to changes in both directions "length of the forecast horizon" and "extremeness of the forecast". Rephrasing the above, the hypothesis tested in this section are:

[25] We found that the NG futures market incorporates 1-day ahead forecasts as a "prior," and when these forecasts are proved to be wrong the following day, the market updates NG future prices in the intraday trading session and aligns NG price to what would be expected from observed temperatures (See Appendix B for details).

[26] "The game of professional investment is intolerably boring and over-exacting to anyone who is entirely exempt from the gambling instinct; whilst he who has it must pay to this propensity the appropriate toll." John Maynard Keynes, The General Theory of Employment, Interest, and Money.

a) the event study Cumulative Abnormal Return (CAR) based on extreme weather forecasts events increases as the forecasting horizon increases;
b) the CAR increases as the threshold for defining a "extreme weather forecast event" increases (i.e. becomes more extreme).

4.2. "Extreme Weather Forecast" risk premium extraction

Typically risk premia are extracted as differentials through *spreading strategies* designed to avoid market directionality. For example, the "Small Minus Big" (SMB) and "High Minus Low" (HML) factors from Fama and French (1992) are differentials in equity portfolios ranked by market capitalization and "book-to-market" ratios, respectively. The Campbell et al. (1997) momentum factor is the differential between the top 20 % "winners" and the bottom 20 % "losers" of stocks ranked on past performance. In the case of listed futures contracts such as NG NYMEX futures, spreading strategies can be easily implemented by any market participant, because the only limitation on entering a long or short position with listed futures is the amount of margin required by the exchange to enter the contract. There are indeed less regulatory and operational limitations on short positions through listed futures, unlike for example for listed stocks.

Spreading strategies are known to extract the commodity curve term premiums (e.g. Boons & Porras Prado, 2019). Spreading strategies on futures curves can take advantage of *Contango*[27] and *Backwardation*[28] of the futures curve. Contango and Backwardation conditions can be exploited through buying and selling futures contracts with different maturities, in a spreading strategy, to capture the price differential between them.

We define the percentage difference between the NG listed future with the closest maturity, the "first month" contract, NG1, and NG listed future with the second closest maturity, "second month" contract, NG2, as "short term contango":

$$short\ term\ Contango_t = \left(\frac{NG2_t - NG1_t}{NG1_t}\right) \times 100 \tag{1}$$

where $NG1_t$ is the 1st month (also known as front-end) NG futures price level and $NG2_t$ is the 2nd month NG futures price level. Fig. 8 shows that NG1 and NG2 futures contracts are usually in a state of "Normal contango", meaning that most of the time the NG2 future price is higher than the NG1 future price. This was the case in 79 % of trading days between 1990 and 2019.

Fig. 9 shows the various quantile regression slopes ("betas") compared to an OLS regression between the *short term contango* and the U.S. daily temperature 1-week forecast differentials from seasonal expectations $temp\_diff_t^{1W\,forecast}$. The short-term contango quantiles react differently to the forecasted temperature differentials from seasonal norms. The lower quantiles of the *short term contango* have a positive coefficient, indicating a direct relationship between short-term contango and forecasted temperature differential from seasonal expectations. In this case, lower temperature forecasts than seasonal expectations imply a stronger *backwardation*, where a nearer delivery future (NG1) is more valuable than a two-month delivery future (NG2) due to a sharp and unexpected rise in demand caused by the exogenous effect of worsening temperature. Conversely, higher temperature forecasts than seasonal expectations during autumn and winter months imply a weaker backwardation[29] or a contango,[30] where first-expiry NG futures (NG1) lose their relative appeal.

The upper quantiles of the short-term contango have a negative coefficient, indicating an inverse relationship with forecasted temperatures. In this case, lower temperature forecasts than seasonal expectations imply a decreasing contango,[31] while higher temperature forecasts than seasonal expectations imply an increasing contango.[32] In an oversupplied environment with a curve already in contango, an exogenous shock, such as colder than expected forecasted temperatures, will decrease the relative cheapness of the first-delivery futures (NG1) compared to the second delivery futures (NG2).

The above observations on the contango/backwardation behavior of the NG futures curve imply that a *spreading strategy* to extract a weather-related risk premium could be implemented as follows within the cold period of the year, e.g. during autumn and winter months:[33]

- in a colder than expected period it would make sense to go *long* on the first expiry future (NG1) while *shorting* the second expiry future (NG2) to take advantage of the *backwardation*, as near-term futures guarantee a close delivery of the physical NG, which is needed now to face the unexpected increase in energy demand led by a temporary unexpected decrease in temperature;
- in a warmer than expected period it would make sense *to short* the first expiry future (NG1) and go *long* on the second expiry future (NG2) to take advantage of the *normal contango of NG* futures.

[27] Contango is a situation where futures prices with longer maturities are higher than futures prices with shorter maturities.
[28] Backwardation is a situation where futures prices with shorter maturities are higher than futures prices with longer maturities.
[29] By "weaker backwardation" we mean a lower positive difference between NG1 and NG2, where backwardation is defined by NG1 price is higher than the NG2 price.
[30] Contango is defined by spot prices, as well as lower maturity futures, having lower prices than longer maturity futures: in our case contango is when NG1 price is lower than the NG2 price.
[31] By "decreasing contango" we mean the positive difference between NG2 and NG1 decreasing, which means that NG1 price increasing respectively to NG2 price. Given NG cost of storage, a decreasing contango often signals an expected increased demand for the physical NG.
[32] "Increasing contango" means NG2 prices are increasing relatively to NG1 price.
[33] Our results show that the identified risk premium has statistically significant effects only during autumn and winter months.

In the analysis that follows, we will focus solely on the extraction of the short-term premium[34] from the NG futures curve through a long position in NG1 and a short position in NG2 futures contracts when temperature is "extremely" cold compared to seasonal norms. Our approach is supported by two key factors. Firstly, the short end of the NG futures curve typically exhibits a state of "normal contango", as shown in Fig. 8: capturing the premium linked to normal contango is a well-established practice in the NG futures market. Secondly, the analysis in Appendix B reveals that only negative weather shocks have a statistically significant impact on NG futures daily returns. Hence, our primary objective is to investigate the impact of deteriorating temperature forecasts on short-term backwardation behavior on the NG futures curve. This specific behavior is crucial for extracting the short-term risk premium associated with extreme cold events from the NG futures market and implies an asymmetric response to "cold" versus "hot" temperature shocks.

In general, extreme temperature events are typically characterized by periods where temperatures reach a level significantly beyond the normal range of temperatures for a given region and period of the year. In our analysis, we define extreme cold weather events as days where the temperature differential from seasonal norms falls below the 10th percentile of historical observations, which aligns with meteorological standards (Kim & Lee 2019, World Meteorological Organization 2023). It is important to note, however, that there are no fixed thresholds that definitively define "extreme temperature events". What is relevant for our analysis is the expectation that the more extreme the temperature event, indicated by a lower temperature quantile, the higher the associated risk premium is likely to be (Fig. 10).

4.3. Extreme weather forecasts risk premium historical performance

We evaluate the effectiveness of a spreading strategy that involves taking a short NG2 position along with a long NG1 position every business day when the "event" temperature differentials (actual/observed and forecasted) from seasonal norms fall below a certain decile threshold. The short NG2 and long NG1 positions are held for a period of 5 business days following the occurrence of the weather event, enabling a comparison of the performance of spreading strategies that maintain open futures positions for the same time frame.

Weather events are defined as the occurrence of a U.S average temperature (as defined in Section 3) below a specific percentile:

$$Event_t : temperature\ below\ Quantile^{th} \equiv temp_{{diff}_{t-1}}^{temp\_type} < Quantile^{th}_{temp} \tag{2}$$

Where:

- $temp_{{diff}_{t-1}}^{temp\_type}$ is the difference from of *temptype* from climatological expectations. Weather events are lagged 1 business day ("t-1") to avoid look-ahead bias.
- *temptype* As shown in Section 2, Table 3, *temptype* can be:
  o $ACTUAL14:00$: Observed temperature at 14:00 EST.
  o $1dforecast$: One day ahead forecasted temperature.
  o $1Wforecast$: One week ahead forecasted temperature.
  o $2Wforecast$: Two weeks ahead forecasted temperature.
- $Quantile^{th}$ is the i-th percentile of observed US average temperature differential from seasonal expectations ($temp_{{diff}_{t-1}}^{ACTUAL14:00}$).

Extreme weather events are defined as temperature below the 10th percentile:

$$Event_t : ExtremeWeatherEvent \equiv temp_{{diff}_{t-1}}^{temp\_type} < Quantile^{10\%}_{temp} \tag{3}$$

Fig. 10 highlights the results over the 30-year period 1990–2019 for different temperature deciles, including the 10th percentile which defines "Extreme Weather Events", while Tables 5–7 details summary statistics and Cumulative Abnormal Returns (CARs) computed accordingly to a standard Even Study methodology (Khotari & Warner, 2006)[35] where "events" are weather events as defined in (2) and extreme weather events are defined in (3).

Table 5 shows that a spreading strategy based on actual (observed) U.S. temperatures does not produce any abnormal return, even for extreme cold conditions (<10 % percentile).

The results presented in Table 6 demonstrate that a spreading NG1-NG2 strategy that relies solely on 1-week forecasted temperatures is capable of extracting the short-term curve premium driven by weather events, even when considering less extreme percentiles such as the 20th and 30th percentiles, in addition to "extreme weather" events below the 10th percentile. Notably, as the percentile of colder-than-expected forecasted temperatures becomes increasingly "extreme" (i.e., lower percentile), as in the case of the 10th percentile, both the CAR and statistical significance of the results tend to increase.

As shown in Table 7, a spreading strategy that exclusively uses 2-week forecasted temperatures is effective in extracting the short-term futures weather-induced premium, particularly when considering "extreme weather" events that fall below the 10th percentile, as well as the 20th percentile: both CARs are significant above 99 % confidence level. Fig. 11 depicts the historical cumulative percentage returns associated with this strategy, spanning the period from 1990 to July 2019, alongside the performance of the S&P500 index for

[34] See the Appendix in Boons and Porras Prado (2019) for a break-down of expected futures returns into spot and term premiums.

[35] See Appendix B for the Event Study methodology and formal definitions of Abnormal Returns and Cumulative Abnormal Returns (CARs).

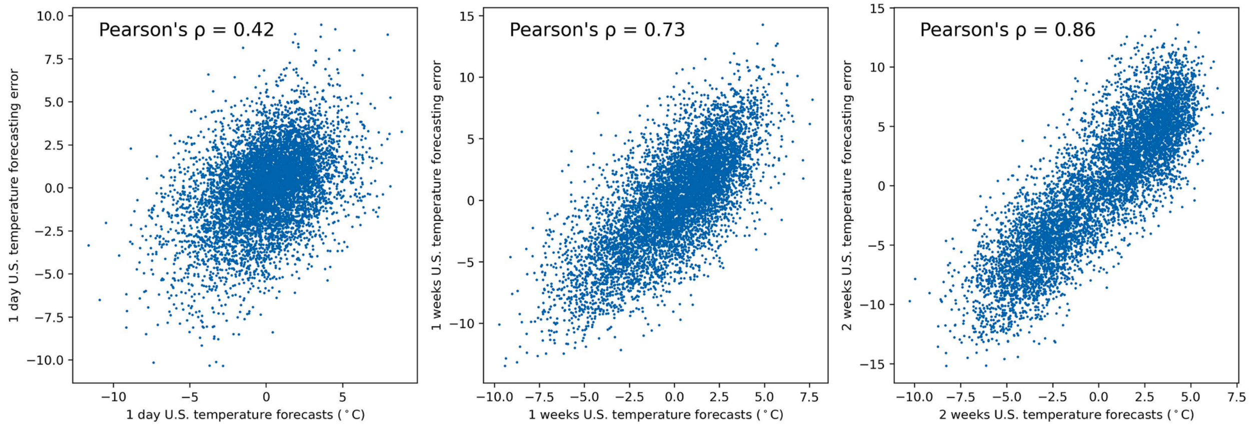

Fig. 7. U.S. daily average temperature forecasts vs ex-post forecasting error, 1990–2019. In the Figure, each dot corresponds to the U.S. forecasts of temperature (weighted average of the 12 biggest U.S. metropolitan areas, adjusted for seasonal expectations (Section 3.1.4) and their ex-post computed forecasting errors. The forecasting errors is measured as the difference between the forecasted temperature and the actual temperature observed ex-post. The figure shows that ex-post forecasting error (y axis) linear correlation to the forecasts (x axis) tend to increase with forecasting horizon: the longer the forecasting horizon, the higher the chance that a forecast significatively different from seasonal expectations will exhibit significant errors. This implies that extreme temperature forecasts tend to entail more risk for market participants when acted upon.

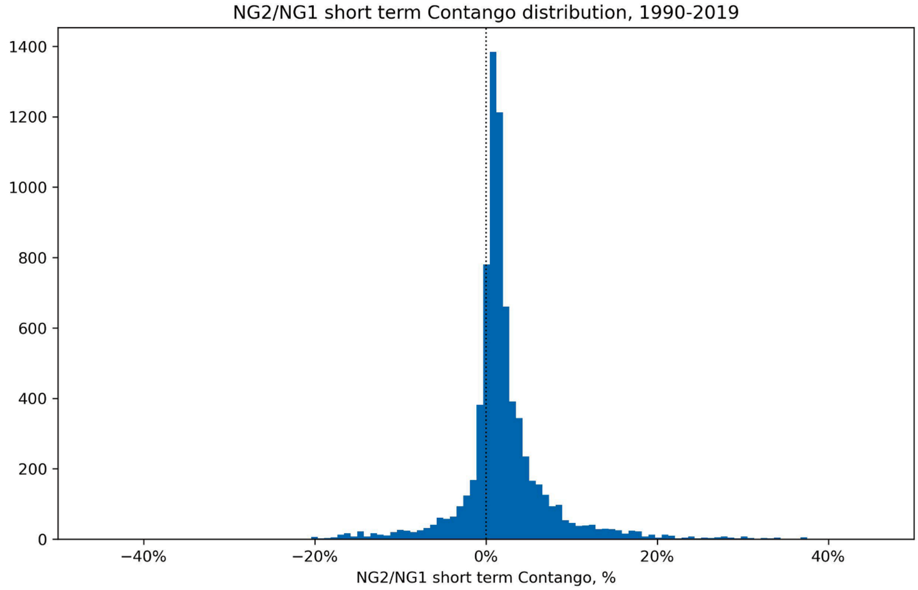

Fig. 8. Short term Contango distribution 1990–2019. Short term contango – distribution, daily data from 4 April 1990 to 31 July 2019. The short end of the NG futures curve tends to be in positive "Normal Contango" most of the time.

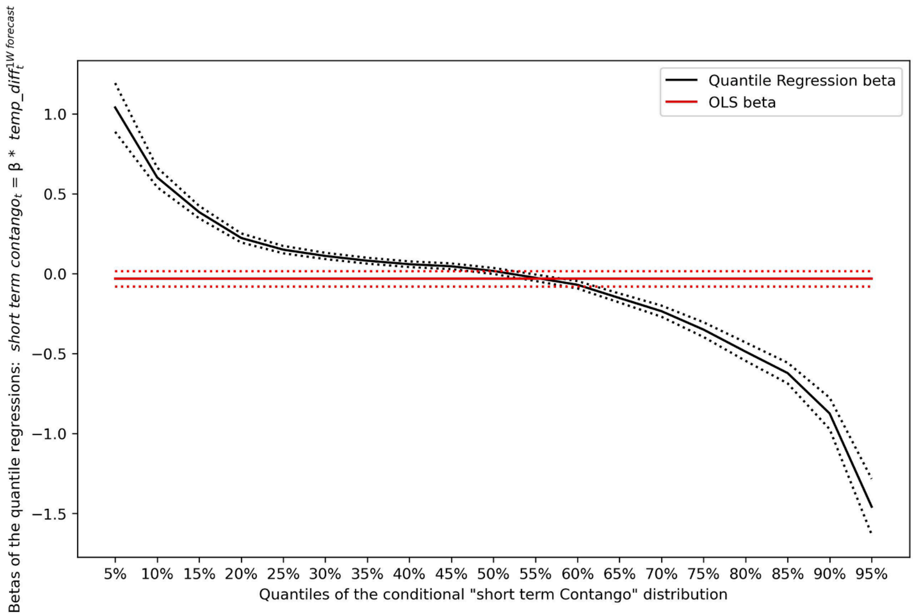

Fig. 9. Short term Contango quantile regression versus U.S. daily average temperatures differentials from norms. Coefficients of the quantile regression between daily NG futures "short term contango" and daily U.S. daily temperature 1-week forecast differentials from seasonal expectations $temp\_diff_t^{1W\,forecast}$. Dotted lines represent 95 % confidence intervals. All quantile regressions are based on daily data from 4 April 1990 to 31 July 2019.

comparison purposes. Notably, we observe that the risk premium associated with "extreme weather events" below the 10th percentile consistently outperforms the S&P500 index, while the risk premium associated with temperature differentials from seasonal norms below the 20th percentile outperforms the S&P500 index to a lesser extent.

With these evidences we confirm our initial hypotheses:

a) the event study CAR based on extreme weather forecasts events does increases monotonically as the forecasting horizon increases, when the CAR is statistically significant;

b) the event study CAR does increases monotonically as the threshold for defining a "extreme weather forecast event" increases (i.e. becomes more extreme), when the CAR is statistically significant.

4.4. Extreme weather forecasts risk premium: comparative performance, seasonality and absence of linear correlation with known NG market fundamentals

As indicated in Table 7, the CAGR of the spreading strategy based on 2-week forecasted temperatures was 12 % over the period from 1990 to 2019. Notably, this rate is significantly higher than the traditionally accepted value of the equity risk premium, which long-term studies have assessed to be approximately 5 % in terms of CAGR (based on data spanning from 1900 to 2015, with a 5 % geometric average return and a 6.5 % arithmetic average return, using nominal yearly rates, as documented in Dimson, Marsh, and Staunton, 2016).

It is noteworthy that during the same period, the Sharpe ratio of the S&P500 index was approximately 0.4,[36] while the risk premia strategies (which utilize only forecasted U.S. temperatures as a trading signal) for "extreme events" (10th percentile) had a Sharpe ratio of 1.3, roughly 3 times higher. A two-sample *t*-test on the difference between them shows that the difference between these two Sharpe ratios is significant above 99.99 %.[37] As a result, we conclude that during the period analyzed (1990–2019), an "extreme weather forecast" risk premium based on the 10th percentile of U.S. temperatures differential from seasonal expectations significantly outperformed the S&P500 index in both absolute and risk-adjusted terms.

Table 8 presents the returns of the extreme weather forecasts risk premia by month of the year, covering the period from 1990 to 2019. This table is complemented by Table 9, which provides the monthly returns of the S&P500 index for comparison purposes.

Table 8 reveals a notable seasonality in the risk premium associated with extreme weather forecasts. It is observed that historically, the majority of the returns from this risk premium occur during the months of October and November. Results in Table 10 indicate absence of statistically significant correlation between the risk premium associated with extreme weather forecasts and the U.S. equity market, represented by the S&P500. Additionally, the correlation of the risk premium does not demonstrate any statistically significant linear relationship with fundamental factors in the NG market.

4.5. Extreme weather forecasts risk premium: risk premium extraction strategy refinements based on the differentiation between "winter" and "summer" months and NG storage levels

In the previous paragraph we analyzed the results of the risk premium extraction via a "naïve" strategy based exclusively on the events of temperature readings materially below seasonal expectations. It is however possible to refine the trading strategy to extract the risk premium by taking into consideration additional triggers such as the season of the year and NG storage levels. The primary results are summarized below, and further details are provided in Appendix D:

- **Differentiation between "winter" and "summer[38]" months:** we highlighted how we observe a strong winter seasonality of the risk premium based on colder than expected events (Table 8): a natural enhancement of this strategy would be to differentiate between winter and summer months. Accordingly, we observe for "winter months" CAR with very high level of statistical significance (above 99 %) for all deciles up to the 50th percentile of temperatures (see Table D.2), which is an improvement compared to the naïve strategy previously discussed (cfr. Table 7). In Table D.2 we can also observe that the CAR is monotonically increasing (when statistically significant) the more the temperatures percentiles decrease, as expected. Overall, the return in term of CAGR of the strategy is lower, while the risk adjusted Sharpe ratio is in the same range for the 10th and 20th percentile of temperatures differential from seasonal norms, while materially higher for 30th, 40th, 50th and 60th percentiles. On the other hand, the same strategy applied in "summer months" shows only the 10th and 20th percentiles having a statistically significant CAR, with however lower CAGR and somehow comparable Sharpe ratios for these statistically significant temperature percentiles.
- **Opposite strategy during "summer" months:** we developed the summer versus winter months analysis further by considering the opposite case as well for summer months: the "S"-shaped curve in the quantile regression between temperatures differentials from seasonal expectations and the contango and backwardation (see Fig. 9) suggests that the opposite strategy may be applied in summer month. Specifically, when temperatures differentials from seasonal expectations rise above a specific threshold, we may want to invert the strategy by going short NG1 and long NG2 in the expectations that the heatwave will reduce NG consumption in the USA. The results for this "opposite" strategy in summer months are highlighted in Fig. D.4 and Table D.4: we do not find an equivalent effect on temperatures rising above the 90th percentile in Table D.4 as we observed for when temperatures descend below the 10th percentile in the naïve strategy (Table 7). Further we do not find CAR monotonicity in function of the temperature

[36] Over the period 1990–2019 the S&P500 index had a CAGR of 7.6% and an annualized Sharpe ratio of 0.43. Note: the S&P500 is an ex-dividend index and the performance of the S&P500 index is therefore underestimated compared to the buy & hold performance of an investor holding the S&P 500 portfolio.

[37] S&P 1990–2019 sample size = 7423, Extreme Cold event 10th percentile, 2-week ahead forecasts, 5 day holding period, sample size = 2326. Sharpe ratios difference (1.3 – 0.43). T-statistic = 31.57 with associated p-value above 99.99%.

[38] We follow here the U.S. natural gas industry terminology where the storage and heating "winter" season is typically defined as November 1 through March 31, while the injection "summer" season is April 1 through October 31.

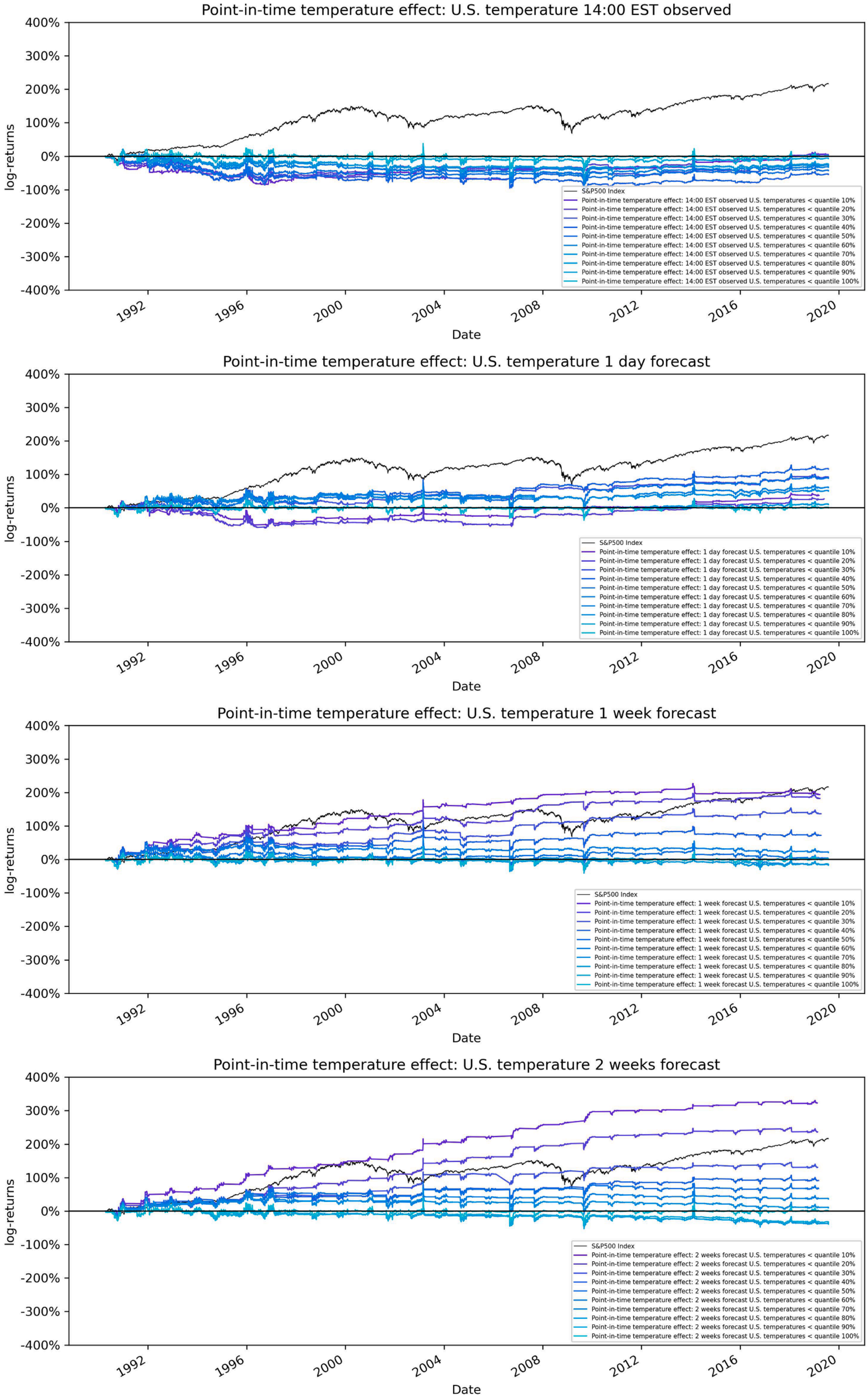

Fig. 10. CAR generated by a point-in-time observed and forecasted temperature differential from seasonal expectations (logarithmic returns). The signal to long NG1 and short NG2 is simply the U.S. average weather difference from seasonal norms to be lower than its nth percentile. Observed U.

S. temperatures do not generate any significant CAR, while 1-week and 2-weeks ahead point-in-time forecasts generate material CARs. Notably, the lower the percentile of temperatures, the higher the CAR. The S&P500 index is shown for comparison purposes. Data from 4 April 1990 to 31 July 2019.

percentiles. The absence of clear results may indicate a more complex dynamic in NG demand during the summer months and further analysis differentiating between residential, commercial, industrial and electric power generation is needed.

- **NG storage levels**: we employ NG storage level information from the U.S. EIA using two alternative approaches differing in the length of historical context applied: a short-term (fast-paced) specification based on a three-month moving average, and a long-term (slow-paced) specification based on a five-year same-week moving average. Results in Tables D.5–D.8 show statistically significant CAR, especially for the lower temperature percentiles but overall lower CAGR compared to the naïve strategy (cfr. Table 7).

The results presented in Appendix D indicate that, overall, the refinements analyzed to the naïve strategy for extracting the risk factor can yield improvements in Sharpe ratios, but do not result in significantly higher CAR. This result is consistent with the notion of a risk premium, which exists to compensate for exposure to non-diversifiable risk rather than to generate excess returns outright.

5. Discussion

As far as we are aware, this is the first comprehensive study in academic literature that has tested the hypothesis of informational efficiency of the U.S. NG futures markets against both realized and forecasted weather conditions and notably, the first study that identified a risk premium associated with extreme weather forecasts. To study the risk premium associated with extreme weather forecasts was first necessary to test more broadly the reaction of NG futures prices in relation to temperature information, we evaluate the hypothesis of the informational efficiency of U.S. NG futures markets against weather conditions by examining the daily and hourly weather data across the entire U.S. for 30 years, starting from the early 1990 s and the complete liberalization of the NG market[39] up to mid-2019. We utilize an event study to define various weather events based on observed and forecasted temperatures (Appendix B). Our analysis shows that NG futures prices do not significantly react to current deterministic weather information based on observed temperatures. However, we find an important exception to this: observed temperatures have an impact on NG futures daily returns when the U.S. average observed temperature at 14:00 EST is not in line with previous day weather forecasts. These findings highlight that the market incorporates 1-day ahead forecasts as a Bayesian “prior,” and when these forecasts are wrong, the market updates NG1 prices in the intraday trading session the following day. This is the first time in the literature that such market behavior has been observed. Appendix B identifies statistically significant NG 1-day market moves related to *probabilistic* weather information (U.S. short term temperature forecasts), particularly when forecasted temperature levels are materially below seasonal expectations. Appendix B also investigates the abnormal returns generated by forecasted temperature daily changes, considering the previous temperature levels. Our findings suggest that the market response to weather information is non-linear and depends on the forecasted temperature levels already priced-in during the previous trading day (t-1). They show that NG futures daily returns respond only to forecasts of temperatures lower than seasonal norms, with the most significant movements observed when the U.S. temperature is predicted to decrease further from levels that are already below seasonal expectations. The results presented in Appendix B indicate that NG futures daily returns are largely impacted by probabilistic weather information, i.e. temperature forecasts. Conversely, deterministic weather information, such as current observed temperature, does not significantly affect NG daily returns except for the case noted above. This applies even in cases where current observed temperatures are extreme.[40] From a market efficiency perspective, our study can be summarized as follows: if NG market is efficient towards observed historical information, and weather forecasts bear a risk of model error given their probabilistic nature, it must be concluded that any abnormal return should stem from market forces rewarding the risk implicit in weather forecasts, reinforcing the principle of “no free lunch”.

From an empirical results perspective, we demonstrated how a straightforward approach utilizing solely weather forecast information (our weather events are defined only by weather information) can successfully capture an “extreme weather forecast” risk premium. Furthermore, we find that this risk premium outperformed the S&P500 in terms of both absolute and risk-adjusted performance over the 30-year period analyzed (1990–2019). It is important to emphasize that the “extreme weather forecast” risk premium can be obtained using solely U.S. weather forecasts, without considering current NG market indicators[41] or any other NG market “fundamentals[42]”. Lastly, we show that the identified risk premium is not correlated to known NG market fundamentals previously identified in the literature.

The significance of a risk premium that depends solely on weather variables, extreme weather forecasts, in this case, is notable. As recent studies suggest that climate change is leading to an increase in extreme weather events, particularly in the form of extreme cold events in the Northern Hemisphere and specifically in the U.S. (Cohen et al., 2018, 2021), our findings offer a quantitative framework

[39] The Henry Hub Natural Gas futures contract on NYMEX has been widely used as a national benchmark price since April 1990. For a summary of the NG regulation history in the U.S., see Hailemariam and Smyth (2019).

[40] This echoes Warren Buffett’s observation that "The investor of today does not profit from yesterday’s growth”.

[41] NG typical “market indicators” are current NG futures price levels, curve shape, price momentum, etc.

[42] NG storage, Crude Light parity, Gulf of Mexico disruptions, etc.

Table 5
Risk premia summary statistics: long NG1/short NG2 spreading strategy based on current observed (point-in-time EST 14:00) temperature levels.

HP 5 day	Quantile %	Quantile °C	CAR	CAGR	Absolute Sharpe ratio	Trading days
$temp\_diff_{t-1}^{actual14:00}$	<10 %	−3.2	7 %	+0%	0.0	2077
	<20 %	−2.0	3 %	+0%	0.0	3578
	<30 %	−1.2	−24 %	−1%	0.1	4754
	<40 %	−0.6	−42 %	−2%	0.1	5606
	<50 %	0.0	−35 %	−1%	0.1	6256
	<60 %	0.6	−28 %	−1%	0.1	6765
	<70 %	1.3	−24 %	−1%	0.0	7091
	<80 %	2.1	−22 %	−1%	0.0	7272
	<90 %	3.3	−6%	−0%	0.0	7382
	100 %		0 %	−0%	0.0	7419

This table reports CARR (Cumulative Abnormal Returns), CAGR (compound annual growth rate), absolute value of the Sharpe ratio and number of active trading days of a strategy consisting of a long position in NG1 and a short position in NG2 held for 5 trading days (one calendar week) when the EST 14:00 observed temperature differentials from seasonal expectations was below a given percentile in t. Back testing period: 1990–2019. Significance levels: (***) 1 %, (**) 5 %, (*) 10% for the test statistic CAR over its standard deviation. The test statistic is assumed to distribute as a standardized Normal (Khotari and Warner, 2006). Simulation based on daily NYMEX futures settlement prices, 4 April 1990 to 31 July 2019, without considering transaction costs and market slippage.

Table 6
Risk premia summary statistics: long NG1/ short NG2 spreading strategy based on forecasted temperature levels, 1-week ahead point-in-time forecast.

HP 5 day	Quantile %	Quantile °C	CAR	CAGR	Absolute Sharpe ratio	Trading days
$temp\_diff_{t-1}^{1Wforecast}$	<10 %	−3.2	595 %	+7 %***	0.9***	1894
	<20 %	−2.0	519 %	+6 %**	0.6**	2916
	<30 %	−1.2	291 %	+5 %*	0.4*	3574
	<40 %	−0.6	105 %	+2 %	0.2	4219
	<50 %	0.0	4 %	+0 %	0.0	4933
	<60 %	0.6	23 %	+1 %	0.0	5678
	<70 %	1.3	−16 %	−1 %	0.0	6376
	<80 %	2.1	−14 %	−1 %	0.0	7017
	<90 %	3.3	1 %	+0 %	0.0	7385
	100 %		0 %	−0 %	0.0	7419

This table reports CAR (Cumulative Abnormal Returns), CAGR (compound annual growth rate), absolute value of the Sharpe ratio and number of active trading days of a strategy consisting of a long position in NG1 and a short position in NG2 held for 5 trading days (one calendar week) when the one week ahead forecasted temperature differentials from seasonal expectations was below a given percentile in t-1. Back testing period: 1990–2019. Significance levels: (***) 1 %, (**) 5 %, (*) 10% for the test statistic CAR over its standard deviation. The test statistic is assumed to distribute as a standardized Normal in the absence of abnormal performance (Khotari and Warner, 2006).

Table 7
Risk premia summary statistics: long NG1/ short NG2 spreading strategy based on forecasted temperature levels, 2-weeks ahead point-in-time forecast.

HP 5 day	Quantile %	Quantile °C	CAR	CAGR	Absolute Sharpe ratio	Trading days
$temp\_diff_{t-1}^{2Wforecast}$	<10 %	−3.2	2425 %	+12 %***	1.3***	2326
	<20 %	−2.0	964 %	+8 %***	0.7***	2970
	<30 %	−1.2	269 %	+5 %**	0.4**	3382
	<40 %	−0.6	149 %	+3 %*	0.2*	3698
	<50 %	0.0	96 %	+2 %	0.2	3989
	<60 %	0.6	43 %	+1 %	0.1	4389
	<70 %	1.3	10 %	+ 0%	0.0	4868
	<80 %	2.1	−32 %	−1%	0.1	5514
	<90 %	3.3	−29 %	−1%	0.1	6647
	100 %		0 %	−0%	0.0	7419

This table reports CAR (Cumulative Abnormal Returns), CAGR (compound annual growth rate), absolute value of the Sharpe ratio and number of active trading days of a strategy consisting of a long position in NG1 and a short position in NG2 held for 5 trading days (one calendar week) when the two week ahead forecasted temperatures differentials from seasonal expectations was below a given percentile in t-1. Back testing period: 1990–2019. Significance levels: (***)1%, (**) 5%, (*) 10% for the test statistic CAR over its standard deviation. The test statistic is typically assumed to distribute as a standardized Normal (Khotari and Warner, 2006).

for assessing the increased costs that NG consumers may face as a result of unpredictable extreme cold events. Indeed, "someone has to pay the price" for bearing the incertitude related to short term extreme weather fluctuations. It is in fact NG final consumers who are paying the market participants the cost of bearing this source of uncertainty by providing liquidity, given that a higher price of the first delivery NG futures tends to affect the price of the physical commodity. Indeed, the NG1 contract holds substantial importance in the

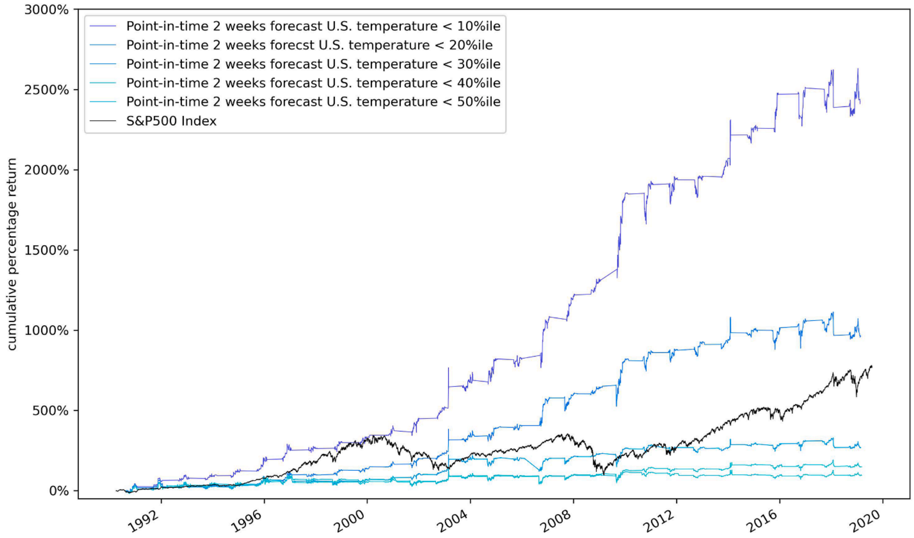

Fig. 11. Cumulative return of risk premium based on 2-week ahead forecasts of colder U.S. weather than seasonally expected, with 5 day holding period (percentage returns). Note: the 10th percentile is the "*extreme weather forecast*" risk premium. The signal to long NG1 and short NG2 is simply the U.S. average weather difference from seasonal norms to be lower than its n^{th} decile. The S&P500 index is shown for comparison purposes. Data from 4 April 1990 to 31 July 2019.

natural gas futures market as it constitutes a significant portion of the overall NG futures open interest.[43] The NG1 contract is particularly noteworthy as it is a physical delivery contract, indicating that the price at its delivery date represents the actual price for physical natural gas delivered at the Henry Hub for the NG1 contract holders. Further, the NG1 price serves as a benchmark for pricing natural gas in physical contracts and spot transactions, and other natural gas futures contracts are evaluated against it. The consequences for the NG market participants and stakeholders are straightforward: the existence of a risk premium tied to extreme weather, coupled with the fact that extreme weather events are on the rise, should lead in the long term to an increased price for NG final consumers, *ceteris paribus*.

This study has several limitations. Firstly, our analysis focuses solely on the U.S. futures market. Although it is currently the largest in the world, it is important to recognize that other regional NG markets, such as the European market, may exhibit different degree of efficiency and/or configurations of risk premia associated with extreme weather forecasts. Secondly, our study examines daily NG futures returns, and it is possible that different patterns may emerge when considering longer-term returns, such as weekly or monthly returns. The dynamics and effects of extreme weather events on NG futures prices may vary over different time horizons, and future research could explore these variations. Furthermore, in order to obtain tractable weather time series for the U.S., and as discussed in Section 2.3, certain decisions may have resulted in simplification and the omission of other effects linked to weather events in the NG futures market. Alternative methodologies,[44] alternative data sources or the use of a more diverse set of weather variables[45] could offer additional insights into the relationship between extreme weather forecasts and NG futures pricing.

Our statistical analysis includes adjustments for False Discovery Rate (FDR) (see Appendix B): after FDR adjustment we do not identify statistically significant impacts of observed temperatures on NG1; it is worth however noting that we observed some instances of significant effect of extreme observed temperatures when considered as a single hypothesis (see Appendix B). The use of techniques to control for FDR remains a topic of ongoing discussion and debate in the field of economics. In the context of our study, the use of FDR control may be relevant to the conclusion drawn from Appendix B, where we find no statistically significant effect of realized temperatures on NG1 daily returns after adjusting for FDR using the Benjamini and Hochberg (1995) procedure, however FDR may hide materially significant effects that in this study we consider as non-statistically significant. In conclusion, while this study provides valuable insights into the risk premia associated with extreme weather forecasts in the U.S. NG futures market, it is important to recognize the limitations inherent in our analysis. Future research should aim to address these limitations by considering other regional markets, exploring different time horizons, refining data methodologies, and further investigating the impact of extreme temperatures on NG futures pricing.

[43] https://www.cmegroup.com/markets/energy/natural-gas/natural-gas.volume.html.

[44] For example, our analysis leverages only "point-in-time" forecasts: as an anonymous Reviewer pointed out "a cold spell and a cold snap have different impacts on storage. A cold spell leads to persistent withdrawals from storage. A cold snap causes a short-term increase in withdrawals but does not lead to sustained pressure on inventories.". Considering the forecasted duration of the weather events may increase the precision of the analysis.

[45] Other weather variables such as rainfall and wind have been documented to have a significant impact on the natural gas market (Shu & Hung, 2009; Liang et al., 2022).

Table 8
Extreme weather forecast risk premium based on 2-week ahead forecasts of colder U.S. weather than seasonally expected, with 5 day holding period, monthly and yearly returns, 1990–2019.

Extreme weather forecast risk premium (<10th percentile 2 weeks ahead forecasts, 5 days)													
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1990				0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−9.1 %	13.6 %	21.2 %	15.0 %	43.9 %
1991	−12.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	8.4 %	0.2 %	30.4 %	−9.7 %	12.0 %
1992	3.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3.9 %	−0.3 %	14.5 %	−9.2 %	10.9 %
1993	−1.8 %	2.7 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−3.6 %	6.4 %	5.6 %	−0.9 %	8.2 %
1994	3.0 %	−0.4 %	−1.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−2.8 %	13.0 %	1.9 %	2.8 %	16.7 %
1995	0.5 %	0.6 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.4 %	5.0 %	7.3 %	13.8 %	30.2 %
1996	8.8 %	−5.6 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	4.5 %	9.1 %	13.5 %	−7.0 %	23.7 %
1997	−0.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−0.1 %	4.8 %	1.4 %	−1.6 %	4.1 %
1998	−1.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3.3 %	0.3 %	6.5 %	1.7 %	10.7 %
1999	−1.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.7 %	6.5 %	0.1 %	0.6 %	6.4 %
2000	1.4 %	2.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.9 %	3.7 %	8.5 %	17.5 %
2001	−6.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−1.4 %	12.5 %	1.0 %	3.6 %	8.8 %
2002	0.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3.6 %	4.7 %	3.9 %	0.9 %	13.6 %
2003	−0.5 %	21.4 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−0.2 %	0.4 %	4.1 %	4.1 %	31.1 %
2004	1.6 %	−4.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	4.9 %	4.6 %	6.4 %	0.5 %	14.0 %
2005	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−1.5 %	−0.8 %	2.8 %	−0.2 %	0.3 %
2006	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−3.2 %	23.7 %	5.7 %	−1.7 %	24.3 %
2007	3.6 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−2.2 %	9.5 %	3.1 %	0.6 %	15.0 %
2008	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2.5 %	0.0 %	3.1 %	0.6 %	7.1 %
2009	0.9 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	17.1 %	9.9 %	4.9 %	2.7 %	40.0 %
2010	0.4 %	−0.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−3.0 %	0.8 %	5.3 %	0.1 %	3.1 %
2011	0.4 %	−0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−5.4 %	5.7 %	2.4 %	−0.3 %	2.2 %
2012	−0.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−4.4 %	5.4 %	0.2 %	0.3 %	0.7 %
2013	−0.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2.0 %	2.3 %	1.1 %	5.5 %
2014	7.2 %	−0.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−0.8 %	2.4 %	0.4 %	8.9 %
2015	0.1 %	−0.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	4.4 %	4.8 %	−0.2 %	9.2 %
2016	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−5.5 %	0.4 %	5.7 %	1.3 %	1.6 %
2017	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−3.6 %	0.7 %	4.4 %	1.9 %	3.1 %
2018	−7.7 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−2.3 %	0.7 %	5.8 %	0.1 %	−3.8 %
2019	−2.2 %	−0.4 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %						−2.6 %
average 1990–2019	−0.2 %	0.4 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	−0.1 %	4.8 %	5.8 %	0.9 %	11.6 %

This table reports CAGR of a spreading strategy consisting of a long position in NG1 and a short position in NG2 held for 5 trading days (one calendar week) when the two week ahead forecasted temperature differentials from seasonal expectations was below the 10th percentile in t-1 from 4 April 1990 to 31 July 2019.

Table 9
S&P500 index monthly and yearly returns, 1990–2019.

S&P 500 index													
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1990				−3.8 %	9.4 %	−0.9 %	−0.5 %	−9.6 %	−5.2 %	−0.7 %	6.1 %	2.5 %	−4.0 %
1991	4.3 %	6.9 %	2.3 %	0.0 %	4.0 %	−4.9 %	4.6 %	2.0 %	−2.0 %	1.2 %	−4.5 %	11.4 %	27.0 %
1992	−2.0 %	1.0 %	−2.2 %	2.9 %	0.1 %	−1.8 %	4.0 %	−2.5 %	0.9 %	0.2 %	3.1 %	1.0 %	4.6 %
1993	0.7 %	1.1 %	1.9 %	−2.6 %	2.3 %	0.1 %	−0.5 %	3.5 %	−1.0 %	2.0 %	−1.3 %	1.0 %	7.2 %
1994	3.3 %	−3.1 %	−4.7 %	1.2 %	1.3 %	−2.7 %	3.2 %	3.9 %	−2.8 %	2.5 %	−4.3 %	1.3 %	−1.6 %
1995	2.5 %	3.7 %	2.8 %	2.9 %	3.7 %	2.2 %	3.3 %	0.0 %	4.1 %	−0.5 %	4.2 %	1.8 %	35.0 %
1996	3.3 %	0.7 %	0.8 %	1.4 %	2.3 %	0.2 %	−4.7 %	1.9 %	5.6 %	2.7 %	7.5 %	−2.2 %	20.8 %
1997	6.3 %	0.6 %	−4.4 %	6.0 %	6.0 %	4.5 %	8.0 %	−5.9 %	5.4 %	−3.5 %	4.6 %	1.6 %	31.8 %
1998	1.0 %	7.2 %	5.1 %	0.9 %	−1.9 %	4.0 %	−1.2 %	−14.9 %	6.4 %	8.2 %	6.1 %	5.8 %	27.4 %
1999	4.2 %	−3.3 %	4.0 %	3.9 %	−2.6 %	5.6 %	−3.3 %	−0.6 %	−2.9 %	6.4 %	2.0 %	5.9 %	20.0 %
2000	−5.2 %	−2.1 %	9.9 %	−3.2 %	−2.2 %	2.5 %	−1.7 %	6.2 %	−5.5 %	−0.5 %	−8.2 %	0.4 %	−10.4 %
2001	3.5 %	−9.4 %	−6.6 %	7.9 %	0.5 %	−2.6 %	−1.1 %	−6.6 %	−8.4 %	1.9 %	7.7 %	0.8 %	−13.3 %
2002	−1.6 %	−2.1 %	3.8 %	−6.3 %	−0.9 %	−7.4 %	−8.1 %	0.5 %	−11.2 %	8.9 %	5.8 %	−6.2 %	−23.8 %
2003	−2.8 %	−1.7 %	0.9 %	8.3 %	5.2 %	1.2 %	1.7 %	1.8 %	−1.2 %	5.6 %	0.7 %	5.2 %	27.1 %
2004	1.8 %	1.3 %	−1.7 %	−1.7 %	1.2 %	1.8 %	−3.5 %	0.2 %	1.0 %	1.4 %	4.0 %	3.3 %	9.2 %
2005	−2.6 %	1.9 %	−2.0 %	−2.1 %	3.1 %	0.0 %	3.7 %	−1.2 %	0.7 %	−1.8 %	3.6 %	−0.1 %	3.1 %
2006	2.6 %	0.1 %	1.1 %	1.3 %	−3.2 %	0.0 %	0.5 %	2.2 %	2.5 %	3.2 %	1.7 %	1.3 %	14.0 %
2007	1.4 %	−2.2 %	1.0 %	4.4 %	3.3 %	−1.8 %	−3.3 %	1.3 %	3.7 %	1.5 %	−4.5 %	−0.9 %	3.6 %
2008	−6.3 %	−3.6 %	−0.6 %	4.9 %	1.1 %	−8.8 %	−1.0 %	1.3 %	−9.3 %	−17.3 %	−7.7 %	0.8 %	−39.2 %
2009	−8.8 %	−11.2 %	8.8 %	9.6 %	5.4 %	0.0 %	7.6 %	3.4 %	3.7 %	−2.0 %	5.9 %	1.8 %	24.1 %
2010	−3.8 %	2.9 %	6.0 %	1.5 %	−8.4 %	−5.5 %	7.0 %	−4.9 %	9.0 %	3.8 %	−0.2 %	6.7 %	13.1 %
2011	2.3 %	3.3 %	−0.1 %	2.9 %	−1.4 %	−1.9 %	−2.2 %	−5.8 %	−7.3 %	11.0 %	−0.5 %	0.9 %	0.0 %
2012	4.5 %	4.2 %	3.2 %	−0.8 %	−6.4 %	4.1 %	1.3 %	2.0 %	2.5 %	−2.0 %	0.3 %	0.7 %	13.7 %
2013	5.2 %	1.1 %	3.7 %	1.9 %	2.1 %	−1.5 %	5.1 %	−3.2 %	3.0 %	4.6 %	2.9 %	2.4 %	30.4 %
2014	−3.6 %	4.4 %	0.7 %	0.6 %	2.2 %	2.0 %	−1.5 %	3.9 %	−1.6 %	2.4 %	2.5 %	−0.4 %	11.7 %
2015	−3.2 %	5.6 %	−1.8 %	0.9 %	1.1 %	−2.2 %	2.0 %	−6.4 %	−2.7 %	8.5 %	0.1 %	−1.8 %	−0.7 %
2016	−5.2 %	−0.4 %	6.8 %	0.3 %	1.6 %	0.1 %	3.6 %	−0.1 %	−0.1 %	−2.0 %	3.5 %	1.9 %	9.8 %
2017	1.8 %	3.8 %	0.0 %	0.9 %	1.2 %	0.5 %	2.0 %	0.1 %	2.0 %	2.3 %	2.9 %	1.0 %	19.9 %
2018	5.8 %	−4.0 %	−2.8 %	0.3 %	2.2 %	0.5 %	3.7 %	3.1 %	0.4 %	−7.1 %	1.8 %	−9.4 %	−6.4 %
2019	8.1 %	3.0 %	1.8 %	4.0 %	−6.7 %	7.1 %	1.3 %						19.4 %
average 1990–2019	0.5 %	0.2 %	1.2 %	1.6 %	0.8 %	−0.3 %	0.9 %	−1.0 %	−0.5 %	1.3 %	1.5 %	1.3 %	7.6 %

This table reports CAGR of the S&P500 index from 4 April 1990 to 31 July 2019.

Table 10
Extreme weather forecast risk premium returns regression with NG market fundamentals and the S&P500.

Independent variables	Regression:$retExtremeWeatherForecasts_t^{2weeksforecasts,10\%}$
Constant	−0.0014***
	(0.0004)
$retCL1_t$	−0.0067
	(0.0147)
$GM_t^{shut-ins}$	0.0000
	(0.0004)
$DOENUST_t^{surprise}$	−0.0160
	(0.0204)
$retS\&P500_t$	−0.0237
	(0.0303)
Adjusted R-squared	0.001
Durbin-Watson stat	2.110
N. observations	2326

This table presents the OLS regression results on the daily log-returns of the Extreme weather forecasts risk premium based on the 10th percentile of U.S. temperatures (differential from seasonal norms) against NG fundamentals and the S&P500 as control variable. $retCL1_t$ is the daily log-return of WTI Crude Light Futures. $GM_t^{shut-ins}$ is the daily amount of Gulf of Mexico production shut-in, in billion cubic feet (see Section 3.2). $DOENUST_t^{surprise}$ is the percentage change of U.S. NG storage compared to its 3-month moving average level in billion cubic feet. Daily data from April 4, 1990 to July 31, 2019. Standard errors are reported in parentheses. ***, **, * denote statistical significance at 1 %, 5 % and 10 % respectively (one tailed test).

In our view, there is a need for further research to delve into the fascinating area of risk premia linked to forecasting errors of exogenous variables. While notable studies like Bessembinder and Lemmon (2002) and Cremers et al. (2008) have explored this topic in the electricity market and corporate credit spreads, the existing literature remains relatively limited. This consideration highlights our study overarching main limitation: although we demonstrate through a statistical model the existence of a novel risk factor stemming from extreme weather forecasts and we propose a potential economic mechanism to explain this premium, we fall short of demonstrating the actual economic mechanism underpinning this new risk factor: further studies on the subject are needed. We hypothesize that a significant portion of risk factors are associated with forecasting errors. Exploring this hypothesis can enhance our understanding of the dynamics and drivers of risk premia in financial markets.

6. Conclusions

In this paper, we present a comprehensive study on the hypothesis of the informational efficiency of U.S. NG futures markets against weather conditions and identify a novel risk premium associated with extreme weather forecasts. By examining 30 years of daily and hourly weather data across the U.S. starting from the early 1990s, we find that NG futures prices do not significantly react to observed temperatures in a daily time frame.[46] On the other hand, we identify statistically significant NG futures 1-day market moves related to weather forecasts, particularly when forecasted temperature levels are below seasonal expectations in autumn and winter months. We demonstrate how this non-linear behavior of short-term weather forecasts can lead to a weather-driven backwardation effect on the short end of the NG futures curve, creating an "*extreme weather forecast*" risk premium. We further show that a classic futures spreading strategy based solely on weather forecast information can successfully capture this unique risk premium, based uniquely on weather data: when the risk premium is extracted for "extreme cold temperatures", defined as the bottom decile of seasonally adjusted temperatures, it outperformed the S&P500 in terms of both absolute and risk-adjusted performance over the 30-year period analyzed.

Our study is important in the context of climate change, which is driving an increase in extreme weather events, particularly in the form of extreme cold events in the Northern Hemisphere and specifically in the U.S. As many recent studies suggest (Cohen et al., 2018; Cohen et al., 2021; Francis et al., 2012; National Academies of Sciences, Engineering, and Medicine, 2016), these events are becoming more frequent, and this has important implications for the energy markets, including the NG market. The identification of a risk premium linked solely to weather variables, such as extreme weather forecasts, is therefore of great importance. Our findings provide a quantitative framework for assessing the increased costs that NG consumers may face as a result of unpredictable extreme cold events, and highlights the need for risk management strategies to mitigate the impact of such events on the energy markets. Overall, this study contributes to a better understanding of the interaction between weather and energy markets, and has important implications for energy market participants, energy policymakers and overall energy risk management.

[46] It is worth noting the exception discussed in Appendix B, when the observed temperatures are considered by market forces to "correct" the previous day wrong weather forecasts incorporated into NG1 market prices.

CRediT authorship contribution statement

Manou Monteux: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Maria Cristina Arcuri:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Gino Gandolfi:** Supervision, Conceptualization. **Stefano Caselli:** Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Professor Giovanni Verga, Professor Marco Riani and Professor Fabrizio Laurini for their valuable feedback.

Appendices A–D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.najef.2025.102494.

Data availability

In order to facilitate the study replication, we provide upon request both the dataset and a Jupyter™ Notebook in Python™ language with a full replication of the study

References

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885–905.
Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy, 7*(2). https://doi.org/10.1155/10.1093/reep/ret016
Bachmeier, L., & Griffin, J. M. (2006). Testing for market integration: crude oil, coal and natural gas. *The Energy Journal, 27*(2), 55–71.
Batten, J. A., Ciner, C., & Lucey, B. (2016). The dynamic linkages between crude oil and natural gas markets. *Energy Economics, 62*, 155–170.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57*(1), 289–300.
Benth, F. E., & Saltyte Benth, J. (2007). The volatility of temperature and pricing of weather derivatives. *Quantitative Finance, 7*(5), 553–561.
Benth, F. E., & Saltyte Benth, J. (2005). Stochastic modelling of temperature variations with a view towards weather derivatives. *Applied Mathematical Finance, 12*(1), 53–85.
Bessembinder, H., & Lemmon, M. L. (2002). Equilibrium pricing and optimal hedging in electricity forward markets. *The Journal of Finance, 57*(3), 1347–1382.
Bigman, D., Goldfarb, D., & Schechtman, E. (1983). Futures market efficiency and the time content of the information sets. *The Journal of Futures Markets, 3*(3), 321–334.
Boons, M., & Porras Prado, M. (2019). Basis-momentum. *The Journal of Finance, 74*(1), 239–279.
Borovkova, S., & Geman, H. (2006). Seasonal and stochastic effects in commodity forward curves. *Review of Derivatives Research, 9*, 167–186.
Bower, R. S., & Bower, N. L. (1985). Weather normalization and natural gas regulation. *Energy Journal, 6*(2), 101–115.
Breyer, S., & MacAvoy, P. W. (1973). The natural gas shortage and the regulation of natural gas producers. *Harvard Law Review, 86*(6), 941–987.
Brigida, M. (2016). The determinants of regime switching in the natural gas and crude oil cointegrating relationship. *Journal of Energy Markets, 9*(3), 17–26.
Brigida, M. (2014). The switching relationship between natural gas and crude oil prices. *Energy Economics, 43*, 48–55.
Brown, S., & Yücel, M. (2008). What drives natural gas? *The Energy Journal, 29*(2), 45–60.
Campbell, J. Y., Lo, A. W., & MacKinley, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.
Campbell, S. D., & Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association, 100*(469), 6–16.
Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics, 81*(1), 27–60.
Cao, M., A. Li, J. Z. Wei. 2003. Weather Derivatives: A New Class of Financial Instruments. *Available at* https://ssrn.com/abstract=1016123 or https://doi.org/10.2139/ssrn.1016123.
Chang, E. C. (1985). Returns to speculators and the theory of normal backwardation. *Journal of Finance, 40*(1), 193–208.
Chiou-Wei, S.-Z., Linn, S. C., & Zhu, Z. (2014). The response of U.S. natural gas futures and spot prices to storage change surprises: Fundamental information and the effect of escalating physical gas production. *Journal of International Money and Finance, 42*, 156–173.
Chiou-Wei, S.-Z., Linn, S. C., & Zhu, Z. (2020). Natural gas price, market fundamentals and hedging effectiveness. *The Quarterly Review of Economics and Finance, 78*, 321–337.
Cohen, J., Pfeiffer, K., & Francis, J. A. (2018). Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nature Communications, 9*, 869.
Cohen, J., Agel, L., Barlow, M., Garfinkel, C. I., & White, I. (2021). Linking Arctic variability and change with extreme winter weather in the United States. *Science, 373*(6559), 1116–1121.
Considine, T. (2000). The impacts of weather variations on energy demand and carbon emissions. *Resource and Energy Economics, 22*, 295–314.
Cremers, K. M., Driessen, J., & Maenhout, P. (2008). Explaining the level of credit spreads: Option-implied jump risk premia in a firm value model. *The Review of Financial Studies, 21*(5), 2209–2242.
Deaves, R., & Krinsky, I. (1992). The behavior of oil futures returns around OPEC conferences. *Journal of Futures Markets, 12*(5), 563–574.
Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics, 4*(3), 66–95.

Demirer, R., & Kutan, A. M. (2010). The behavior of crude oil spot and futures prices around OPEC and SPR announcements: An event study perspective. *Energy Economics, 32*(6), 1467–1476.
Easton, Z. M., Kleinman, P. J., Buda, A. R., Goering, D., Emberston, N., Reed, S., Drohan, P. J., Walter, M. T., Guinan, P., Lory, J. A., & Sommerlot, A. R. (2017). Short-term forecasting tools for agricultural nutrient management. *Journal of Environmental Quality, 46*(6), 1257–1269.
Elkhafif, M. A. T. (1996). An iterative approach for weather-correcting energy consumption data. *Energy Economics, 18*, 211–220.
Enger, I. (1959). Optimum length of record for climatological estimates of temperature. *Journal of Geophysical Research, 64*(7), 779–787.
Erdős, P. (2012). Have oil and gas prices got separated? *Energy Policy, 49*(C), 707–718.
Ergen, I., & Rizvanoghlu, I. (2016). Asymmetric impacts of fundamentals on the natural gas futures volatility: An augmented GARCH approach. *Energy Economics, 56*, 64–74.
Fama, E. F. (1970a). Session topic: Stock market price behavior session chairman: Burton G. Malkiel efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417.
Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review, 10*(1), 1–21.
Fama, E. F., & French, K. R. (1988a). Business cycles and the behavior of metals prices. *The Journal of Finance, 43*, 1075–1093.
Fama, E. F., & French, K. R. (1988b). Permanent and temporary components of stock prices. *Journal of Political Economy, 96*(2), 246–273.
Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums and the theory of storage. *The Journal of Business, 60*, 55–73.
Fama, E. F. (1970b). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417.
Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427–465.
Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49*, 283–306.
Fama, E. F. (1965). The behavior of stock market prices. *The Journal of Business, 38*, 34–105.
Fleming, J., Kirby, C., & Ostdiek, B. (2006). Information, trading, and volatility: Evidence from weather-sensitive markets. *The Journal of Finance, 61*(6), 2899–2930.
Francis, J. A., Stephen, J., & Vavrus, S. J. (2012). Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters, 39*(6), Article L06801.
Gebre-Mariam, Y. K. (2011). Testing for unit roots, causality, cointegration, and efficiency: The case of the northwest US natural gas market. *Energy, 36*, 3489–3500.
Geman, H., & Ohana, S. (2009). Forward curves, scarcity and price volatility in oil and natural gas markets. *Energy Economics, 31*, 576–585.
Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance, 17*(1), 35–105.
Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal, 62*(2), 47–68.
Hailemariam, A., & Smyth, R. (2019). What drives volatility in natural gas prices? *Energy Economics, 80*, 731–742.
Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneu, T. J., Jr., Zhu, Y., & Lapenta, W. (2013). NOAA's second-generation global medium-range ensemble reforecast dataset. *Bulletin of the American Meteorological Society, 94*(10), 1553–1565.
Hartley, P., Medlock, K., & Rosthal, J. E. (2008). The relationship of natural gas to oil prices. *Energy Journal, 29*, 47–65.
Hu, J.-W., Hu, Y.-C., & Lin, C.-Y. (2014). Effect of temperature shock and inventory surprises on natural gas and heating oil futures returns. *The Scientific World Journal, 457636*. https://doi.org/10.1155/2014/457636
Huurman, C., Ravazzolo, F., & Zhou, C. (2012). The power of weather. *Computational Statistics & Data Analysis, 56*(11), 3793–3807.
International Energy Agency (IEA). 2022. Gas Market Report, Q4-2022. *Available at* https://www.iea.org/reports/gas-market-report-q4-2022.
Kellard, N., Newbold, P., Rayner, T., & Ennew, C. (1999). The relative efficiency of commodity futures markets. *The Journal of Futures Markets, 19*(4), 413–432.
Khotari S. P., and J. Warner. 2006. Econometrics of Event Studies. In: Eckbo, B. E. (Ed.), Handbook of Corporate Finance: Empirical Corporate Finance, Handbooks in Finance 1: 3-36.
Kim, Y., & Lee, S. (2019). Trends of extreme cold events in the central regions of Korea and their influence on the heating energy demand. *Weather and Climate Extremes, 24*(100199). https://doi.org/10.1016/j.wace.2019.100199
Kotz, M., Wenz, L., Stechemesser, A., Kalkuhl, M., & Levermann, A. (2021). Day-to-day temperature variability reduces economic growth. *Nature Climate Change, 11* (4), 319–325.
Liang, C., Xia, Z., Lai, X., & Wang, L. (2022). Natural gas volatility prediction: Fresh evidence from extreme weather and extended GARCH-MIDAS-ES model. *Energy Economics, 116*(C). https://doi.org/10.1016/j.eneco.2022.106437
Lin, B., & Li, J. (2015). The spillover effects across natural gas and oil markets: Based on the VEC–MGARCH framework. *Applied Energy, 155*(1), 229–241.
Liu, Y., Dong, H., & Failler, P. (2019). The oil market reactions to OPEC's announcements. *Energies, 12*(17), 3238. https://doi.org/10.3390/en12173238
Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Science, 20*, 130–141.
MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature, 35*(1), 13–39.
Martinez, B., & Torro, H. (2015). European natural gas seasonal effects on futures hedging. *Energy Economics, 50*, 154–168.
McGuirk, M., Shuford, S., Peterson, T. C., & Pisano, P. (2009). Weather and Climate Change Implications for Surface Transportation in the USA. Available at *World Meteorological Organization Bulletin, 58*(2) https://library.wmo.int/doc_num.php?explnum_id=6938.
Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics, 29*, 46–63.
National Academies of Sciences, Engineering, and Medicine. (2016). *Attribution of Extreme Weather Events in the Context of Climate Change*. Washington, DC: The National Academies Press.
Nick, S., & Thoenes, S. (2014). What drives natural gas prices? A structured VaR approach. *Energy Economics, 45*, 517–527.
Ramberg, D. J., & Parsons, J. E. (2012). The weak tie between natural gas and oil prices. *Energy Journal, 33*, 13–35.
Richman, M. B., & Lamb, P. J. (1985). Climatic pattern analysis of three- and seven-day summer rainfall in the Central United States: Some methodological considerations and a regionalization. *Journal of Applied Meteorology and Climatology, 24*, 1325–1343.
Roll, R. (1984). Orange juice and weather. *American Economic Review, 74*, 861–880.
Samuelson, P. A. (1965a). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review, 6*, 41–69.
Samuelson, P. A. (1965b). Rational theory of warrant pricing. *Industrial Management Review, 6*(2), 13–39.
Scott, C. L., & Zhu, Z. (2004). Natural gas prices and the gas storage report: Public news and volatility in energy future markets. *The Journal of Futures Markets, 24*(3), 283–313.
Shao, C., Bhar, R., & Colwell, D. B. (2015). A multi-factor model with time-varying and seasonal risk premiums for the natural gas market. *Energy Economics, 50*, 207–214.
Siliverstovs, B., L'Hégaret, G., Neumann, A., & Hirschhausen, C. (2005). International market integration for natural gas? A cointegration analysis of prices in Europe, North America and Japan. *Energy Economics, 27*(4), 603–615.
Song, J. H., M. L. De Prado, H. Simon, and K. Wu. 2015. Intraday Patterns in Natural Gas Futures: Extracting Signals from High-Frequency Trading Data. https://doi.org/10.2139/ssrn.2657224.
Till, H. 2018. An Update on Weather Fear Premia Trades, Economic Anthropology. Available at http://premiacap.com/publications/EDHEC-Risk%20an_update_on_weather_fear_premia_trades%20Dec%202018.pdf.
Till, H. (2008). Case studies and risk management in commodity derivatives trading. In H. Geman (Ed.), *Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy* (pp. 255–291). Chichester, UK: John Wiley & Sons Ltd.
Till, H., & Di Tomasso, J. (2000). Active "commodities-based" investing. *The Journal of Alternative Investments, 3*(1), 70–80.
Till, H., & Eagleeye, J. (2005). Commodities – Active strategies for enhanced return. *Journal of Wealth Management, 8*(2), 42–61.
Todorova, M. I. (2004). Modeling energy commodity futures. is seasonality part of it? *The Journal of Alternative Investments, 7*(2), 10–32.

Turvey, C. G. (2001). Weather derivatives for specific event risks in agriculture. *Applied Economic Perspectives and Policy, 23*(2), 333–351.
Wang, T. T., Zhang, D., & Broadstock, D. C. (2019). Financialization, fundamentals, and the time-varying determinants of US natural gas prices. *Energy Economics, 80*, 707–719.
World Meteorological Organization, 2023. Guidelines on the Definition and Characterization of Extreme Weather and Climate Events. Available at https://public.wmo.int/en/resources/library/guidelines-definition-and-characterization-of-extreme-weather-and-climate-events.