Predicting Weather: Forecasting using Futures Contracts and Information Markets

by

Jennifer Yu-Tin Tsao

An essay submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

August 2015

Abstract

In this essay, we propose the use of information markets to predict annual snowfall accumulation in New York City. Our goal is to examine whether the aggregation of private information amongst individuals has the ability to forecast a phenomenon that is long-accepted to be exogenous: weather. Specifically, futures contracts data were used as the representative information market. Modelling snowfall as count data, the variables Closing Price, Open Interest, and variations of the two were used as regressors in Negative Binomial models.

In the latter half of the essay, we compare our futures market forecast of snowfall with meteorological forecasts.

Acknowledgements

First and foremost, I would like to thank Dr. Frank Milne for his genuine help and support. This essay topic is by no means commonly-tread territory[1] and would not have been half as decent without his insight and guidance.

I would like to thank and acknowledge the use of data from the following sources: snowfall data from the National Centers for Environmental Information, *Intrade* data from the site of Dr. Panos Ipeirotis of New York University, and meteorological forecasts from the *Dark Sky* API.

I would like to thank my family for their continued support in my endeavors. I would like to thank the friends I met here at QED for humouring me whenever I made wonderfully dorky economics-related jokes.[2]

I would like to thank A for sharing every lunch and dinner with me in the crummy yet charming cafeteria of Mac-Corry.

Lastly, I would like to thank my wrists for not getting carpal tunnel this year. This essay could not have been penned without them. We did it, guys!

[1] As readers will see from the scant length of my literature review!
[2] And for helping my work and I 'shine bright like a SAS diamond.'

Table of Contents

1. Introduction

> "But who wants to be foretold the weather? It is bad enough when it comes, without our having the misery of knowing about it beforehand."
> — Jerome K. Jerome, *Three Men in a Boat*

As we know from personal experience, weather can be unpredictable. Albeit the technologically sophisticated tools meteorologists have at their disposals, generally one can recall instances of when the weatherman 'got it wrong.' Weather, for good reason, is typically treated as exogenous in academic literature.

Weather lures great curiosity because it is essentially an all-powerful, age-old force that distorts our intended outcomes. A restaurant that loses two customers tonight due to heavy snow may not be able to sell two *extra* meals to make up for loss sales next week. Agents experience actual economic losses due to bad weather; therefore, without a doubt, there is tremendous value in being able to forecast weather accurately. Whole industries, particularly those who are ill-abled to contain inventory, are the most sensitive to the burden of weather fluctuations. To be able to simply *know* the weather in advance results in more control in inventory management. From as short as twelve hours in advance to an entire week, possessing accurate weather forecasts can translate into economic gains.

In this paper, we suggest a creative new method for predicting snowfall by way of using information markets. Otherwise known as 'prediction markets,' they are designed to tease out privately-held information from market participants. The most curious and spectacular aspect of information markets is that they are able to aggregate private information together. Taken together, the bits and pieces of proprietary information together may be powerful enough to predict future outcomes. We appeal, in a sense, to the old adage that is the 'wisdom of crowds.'

In this paper, we explore the ability and effectiveness of a weather futures prediction market in forecasting the amount of accumulated snowfall in Central Park, New York.

1.1 Motivation: Weather Fluctuations are Costly

For certain industries such as the fashion and restaurant businesses where inventory management is very tight, weather fluctuations can incur profit loss. The main problem of these industries is that there is no balancing effect.[3] When a heavy snow storm hits during the night, some people will find their marginal cost of eating out augmented due to the bad weather and instead, decide to stay home. While a furniture seller will likely be able to sell a couch this season no matter what weather occurs today, a restaurant that loses a family of four as customers on a Tuesday may not necessarily make up their business tomorrow, or next week, when the weather is better. People's preferences can change in the short-run. Another angle of thinking about the balancing effect problem is that weather fluctuations can lessen sales volume. Rather than omitting discrete sales like a meal, weather can also lessen sales of products whose nature is a continuous variable. For instance, hydroelectric power companies will generate less electricity – and therefore, *sell* less electricity – when there is less rainfall.

1.2 Weather Derivatives: Curious and Atypical

The examination of the ability of information markets to predict was precisely chosen to be carried out with weather derivatives because they have special characteristics unlike traditional, vanilla contracts. The most obvious difference is that the underlying commodity of a weathers futures is not traded in the spot market.[4] This is important for our purposes because it eliminates a secondary venue for insurance, thereby focusing all the information signals of what the market believes will happen into futures contracts.

---

[3] Jewson and Anders give other examples of businesses that can suffer from a lack of balancing effect: An amusement park will receive fewer visitors when it rains; a natural gas supply company will sell less gas in warmer winters.
[4] For example, one cannot 'buy that it is raining' like they can barrels of gas for immediate delivery.

For instance, a person requiring soybeans two months from now, but who suspects prices will rise in the meantime, has two choices: they can either purchase the soybeans they require today on the spot market, or enter into a futures contract for delivery two months later, thereby locking in the price. The private information that soybean prices will rise can be diverted between these two outlets, but this works against our primary goal because price signals from the spot market is ambiguous: people may be purchasing soybeans simply because they require some to use today, and not necessarily because they believe prices will be rising. This ambiguity problem is eliminated by the non-existence of weather spot markets.

Secondly, weather is specific to its particular geographical location. A particular weather contract will most likely only be traded by the people who are affected by it, especially if it was bought as insurance in the first place.

Lastly, and most importantly for our purposes, there is no issue of moral hazard. One cannot influence a weather contract's underlying outcome by any mortal methods. This point eliminates one of the common arguments against the power of prediction markets, that market participants may inherently be biasing the outcome for economic gains. A person with a long position in a rainfall contract will not be able to make it more likely to rain in order to increase the payout of their contracts!

2. Literature Review

There is very little literature on the prediction of weather using information markets. As far as we know, our study is the first to examine the effectiveness of information markets on an accumulated weather figure such as snowfall.

Earlier related literature stemmed from the field of futures and options. In Roll (1984), it was reasoned that futures prices should take anticipated weather changes – the *predictable* portion of weather – from publicly published meteorological forecasts into account. Roll theorized that price changes in futures prices are solely due to unanticipated, exogenous weather shocks. The author was able to show that orange juice prices were a significantly better predictor of evening temperature forecast error than the forecast error made on the previous day.[5] Roll found that around 90% of the variability in temperature was accounted for by meteorological forecasts. As for the remaining 10%, orange juice prices were able to predict a significant portion of it, beyond what the meteorological forecast was able to do. This essentially means that futures prices are able to significantly predict the variability in weather.

Aside from temperature, Roll also attempted to predict rainfall levels using the same set of futures data. Being a type of precipitation, this was of especial interest to us; however, Roll found that orange juice futures had no significant predicative power for rainfall.

Simmons and Sutter (2011) conducted regression analysis on tornadoes using dataset of over 50,000 tornadoes between 1957 and 2007. A variety of regressors were used, including the strength of the tornado, time when casualties occurred, and various demographic variables.[6] Two independent variables were examined: fatalities and injuries. A Poisson model was used to model fatalities and the Negative Binomial for injuries, due to evidence of overdispersion. This research bares great similarity to this essay because despite the great number of observations, most tornadoes do no kill or injure, thereby leaving Simmons and Sutter with a dataset of only 1,300 tornadoes with which to run regressions. In other words, this dataset contains an excess number of zeroes, much like our snowfall data.

[5] Roll defined forecast error, in this context, to be the difference between the temperature forecast issued by the National Weather Service and the actual outcome.
[6] Examples of demographic variables used include the population density in an area (the more people in an area, the more likely that higher amounts of people are hit by the tornado) and *income* (research has shown that people consider safety a luxury good; wealthier people will purchase weather ratios and install tornado shelters, etc.).

The above two citations represent the scant existing literature we found directly related to the topic of this essay; however, it is possible to draw lessons from other, related research. Most of the existing literature concerned market design, answering questions regarding best practices for policy analysis, for small and illiquid markets, or to prevent foul play within the markets. Keeping this in mind, we shall now discuss what we felt were the most pertinent for our purposes.

Wolfers and Zitzewitz (2004) theorize that in a truly efficient information market, where payoffs are tied to eventual outcomes that will occur in the futures, the most accurate predictor of the outcome will be the market price.[7] As part of the definition, they further assert that no combination of polls or other information will be more accurate than the prediction given by the market.

Experiments have been conducted to show that information does eventually aggregate in market equilibrium. McKelvey and Ordeshook (1985) demonstrate in a laboratory experiment that elections where voters who each had private information about candidates as well as public signals via polling eventually saw information aggregate to form one majority opinion. Taking this into account, we would expect the futures prices in our proposed study to converge over time towards an outcome indicating one clear winner.

In practice, Galebach, Pennock, Servan-Schreiber, and Wolfers (2004) designed an experiment to show that prediction markets based on play-money can be just as accurate as those based on actual money. Their result demonstrates that platforms without the goal of monetary gain can be accurate in predictions. We mention this in order lend some credibility to our results, which are derived from futures contracts of miniscule monetary value.

[7] *Information Markets: A New Way of Making Decisions*, Ch. 1.

3. Data

3.1 Intrade Platform

For our representative futures market, data was used from the now defunct *Intrade.com*.[8] Intrade was a web-based trading exchange whose members traded contracts with one another based on what traders think would occur. It is essentially a simplified futures exchange, except available contracts are based off of events that stock exchanges do not offer. For instance, one can take a position on who they think will be the next winner of Best Picture at the Academy Awards, or the next Democratic nominee in a presidential election.

Since contracts are quite small, without doubt it is possible that Intrade serves as a vehicle for speculators; however, we believe that the Intrade market also attracts *informed*, *purposeful* traders primarily because it can serve as an insurance medium against real events. For instance, it would be difficult to find a contract at a financial institution to hedge the risk for something as specific as President Bush's approval rating passing 60% by December 31, 2006.[9] On the other hand, such contracts *do* exist on Intrade, meaning that agents who stand to gain or lose based on this outcome will flock to Intrade as a form of insurance. This necessity makes Intrade a suitable prediction market.[10]

3.2 Snowfall Futures Contracts

As a proxy for how much snow the market thinks will fall in New York City, we used historical trading data on futures contracts based off of snowfall accumulation in Central Park between November 1st to March 31st of each year.[11] For every winter season,

---

[8] Intrade.com has been defunct as of December 2014.
[9] We would also expect finding a counterparty to be difficult, in this case, for other reasons.

[11] We have data until April 1st of each year, but most futures contracts did not fluctuate in April, and in some cases, contracts ended before April.

contracts ranging from accumulations of 10+ inches, 20+ inches, to 70+ inches are sold. Intrade specifically writes the contracts to pertain to mutually exclusive outcomes. For example, Barack Obama being elected as President in 2012 is a binary outcome - he is either elected or he is not. This exclusion feature ensures that market participants are able to play positions as accurately as possible to what they think or know will be true.

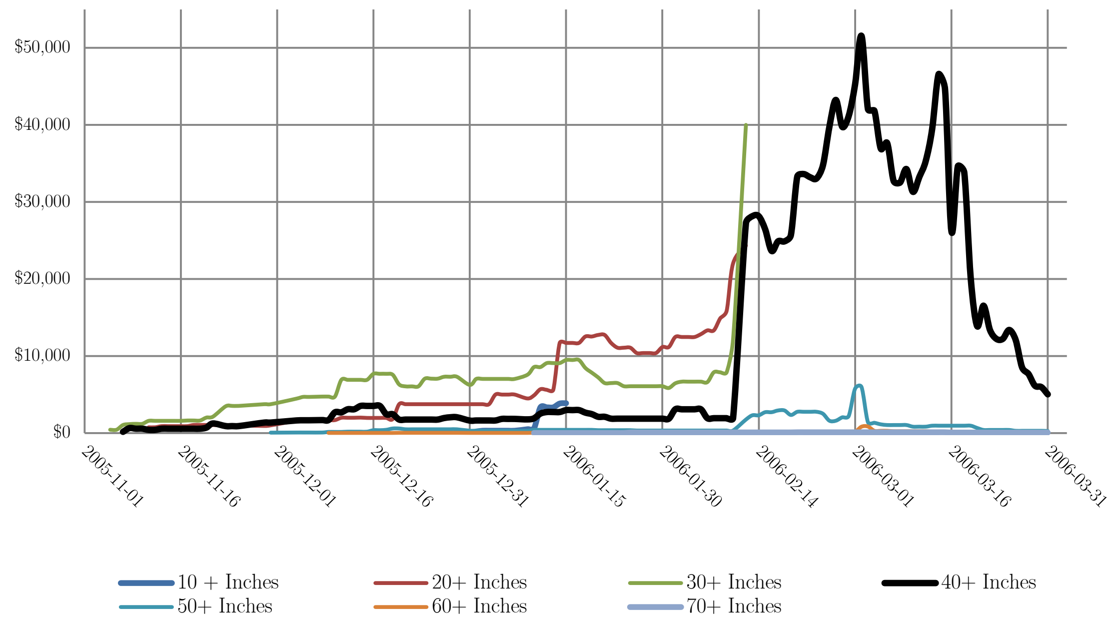
In terms of snowfall, a contract will settle at $0 if snowfall is less than the number of inches specified in the contract. A contract will settle at $10 if snowfall is equal to or more than the number of inches specified in the contract.[12] For instance, if accumulated snowfall measured from November 1st to April 30th is 37.1 inches, then the 40+ and 50+ inches contracts would settle at $0. Traders who took long positions on the above two contracts will lose the amount equal to their initial invested principal. The 30+ inches contract – in this case, the winning contract – will settle at $10 per contract. The 10+ and 20+ inches contracts are technically supposed to pay off according to the contract rules, but in every case we examined from 2003 to 2013, contracts satisfying the correct amount of snow, but were further away from the actual amount of snowfall, were retired as soon as the accumulated snowfall entered the next, ascending category of contracts.

As an example, below lies contract values for futures contracts trading in the 2005 to 2006 winter season, where it ultimately snowed an accumulation of 40.0 inches. The 'Value' of a contract on a given day is henceforth defined to be the Closing Price multiplied by the Open Interest. Trading hours lasted from 3:45 AM to 3:00 AM the following morning in Eastern Time each day. This is equivalent to 8:45 AM to 8:00AM in Ireland, where Intrade was operating.

Figure 1 *Value of Futures Contracts Trading in the 2005 – 06 Winter Season*

[12] Please see Table 1 in the Appendix for contract rules.

2005-06 Value (Winner: 40+ Inches)



3.3 Actual Snow Fall Data

For actual snow fall levels, we used data from the National Centers for Environmental Information gathered between the years 1890 to 2004 at the weather station in Central Park.[13] Observations occurring on February 29th were taken out of leap years because they do not occur every year. We wanted to render the data set to include observations from the *same* dates across different years.

By industry standards set by the National Weather Service (NWS), daily snowfall is measured at midnight each day of the local time zone. In order for snowfall to be consistently measured over time, NWS requires that stations record the maximum level

13 New York Central Park Obs Belvedere Tower, NY.

of snowfall that has fallen since the last record.[14] Measurement is taken to the nearest 0.1 inch.

Like much of weather-related data, there is no hope of stationarity. From historical observations over the last sixty years,[15] the majority of days exhibited no snowfall. It is clear that simply first differencing, or any higher degrees of differencing however number of times, will not achieve stationarity since most values are sandwiched between zeroes. The dataset is not fit for time series techniques.

3.4 Meteorological Forecasts

We obtained forecasts by meteorologists using the API named *Forecast* for the popular mobile application, *Dark Sky*.[16] Dark Sky contains a wealth of weather information gathered by professionals working at government weather stations. The API allows users to query most locations around the world and returns weather data such as current conditions, hour-by-hour forecasts, and even minute-by-minute forecasts for the next day. For historical snowfall forecasts, only the forecasts for the snowfall in the next 24 hours are retained in the database. Historical day-to-day forecasts were not kept, which is unfortunate because it limits our scope to examining the relative effectiveness between meteorological and futures contracts' forecasts up to 24 hours ahead.

[14] NWS defines snowfall as the accumulation of new snow and ice since the last observation, prior to melting or settling. Snowfall is measured as soon as it stops snowing. Taken from the *Snow Measurement Guidelines for National Weather Service Surface Observing Programs* (September 2013).

[15] Please see Tables 2a to 2h in the Appendix.

[16] API, or Application Programming Interface, is a set of data structures, functions, tools, and objects used in software programming. It was fortunate for us that Dark Sky's API was hosted online and access was free of charge.

4. Model

4.1 Initial Assessment

We first start off with a simple, sample observation: Between the years of 2003 to 2008 during which Intrade offered contracts on Central Park's annual accumulated snowfall, the markets were able to successfully predict the snowfall outcome in most years well ahead of contract termination. Evidently, it is not farfetched to presume that futures contracts have some degree of capability to predict.

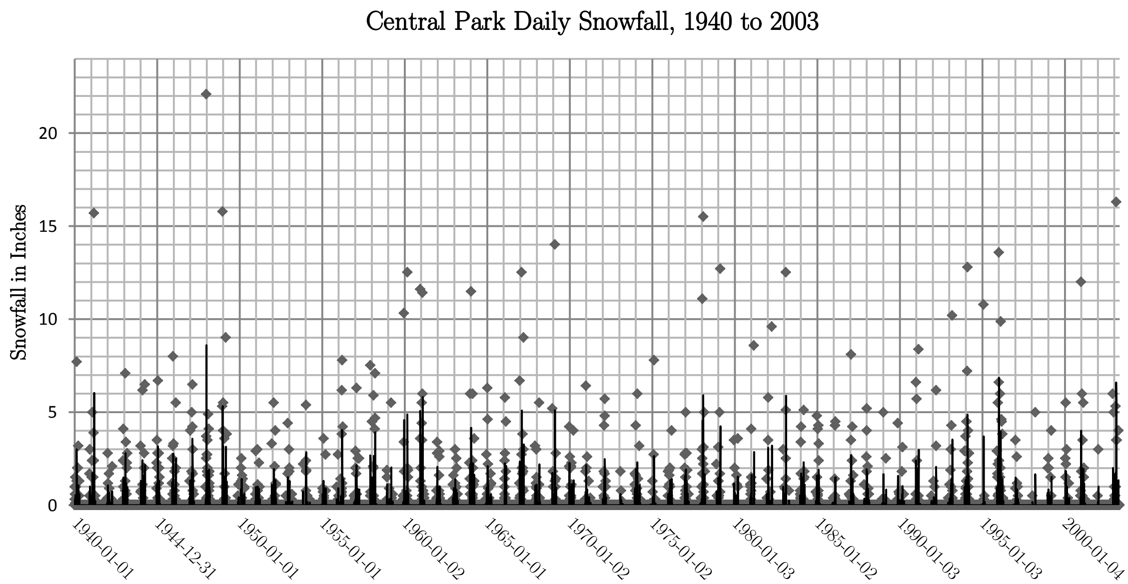
Table 2 *Sample Prediction Ability of Futures Contracts*

Year	*2003-04*	*2004-05*	*2005-06*	*2006-07*	*2007-08*
Accumulated Snow Fall (inches)	42.6	41.1	40.0	12.4	11.9
VALUE					
Did the winning contract predict the correct snowfall accumulation?	Yes	Yes	Yes	Yes	No
Since when?	January 20th, 2004	March 15th, 2005	February 13th, 2006	January 10th, 2007	Markets predicted that the 20+ inches contract would win.
PRICE					
Did the winning contract predict the correct snowfall accumulation?	Yes	Yes	Yes	No	Yes
Since when?	January 13th, 2004	March 15th, 2005	February 22nd, 2006	Markets predicted that the 20+ inches contract would win.*	December 13th, 2007

**March 2008 futures data were cut off for unexplained reasons.*

Modelling snowfall required some creativity on our part due to lacking existing literature. In order to obtain an idea of what we were dealing with, we firstly started with graphing historical daily Central Park snowfall levels in inches. Below, we graph the snowfall data from January 1st, 1940 to April 30th, 2003.[17] There does not appear to be a drift in any particular direction over the years.

Figure 2



From observation, spikes in snowfall mainly occur between the months of December and early March of the following year.[18] Though infrequent, very few spurts of snow occur in November or April. For this reason, we will only be analyzing snowfall between December 1st and March 31st of each season. Additionally, the variables of most futures contracts which terminated before April do not move at all; henceforth, we will only be analyzing contract movements from November 1st to March 31st of each year. November futures were included in our analysis because anticipations for December snowfall will have shown up, theoretically, in November contract movements.

Next, we examine the occurrences of different levels of snowfall over each day in order to see whether certain days or weeks during the winter season have a penchant for snowing

[17] Data from prior to 1940 contained missing values and were thus omitted.
[18] Figure 2: The jagged black lines represent the 3-day moving average. The 3-day period was chosen arbitrarily because most occurrence of snowfall seems to last no more than 3 consecutive days.

relatively more than others. Looking at the tabulated figures in Tables 2a to 2i in the Appendix, we see that snow falls more often and in greater amounts during January and February.

'More than 2.0 inches of snowfall' was chosen to be its own separate category because *among* the days that *did* snow from December 1st to March 31st each year between the years 1940 and 2003, the average of snowfall level was 2.0 inches. In other words, average snowfall conditioned on snow occurring is 2.0 inches per day. Any occurrence of more than 2.0 inches in a day can therefore be interpreted as an 'above average' level of snowfall.

Table 3 *Summary of Snowfall Frequencies, December 1st, 1940 to March 31st, 1999*

	Frequency of Days with Snow	Frequency of Days with Above Average Snowfall
December	123	46
January	181	54
February	147	69
March	100	45

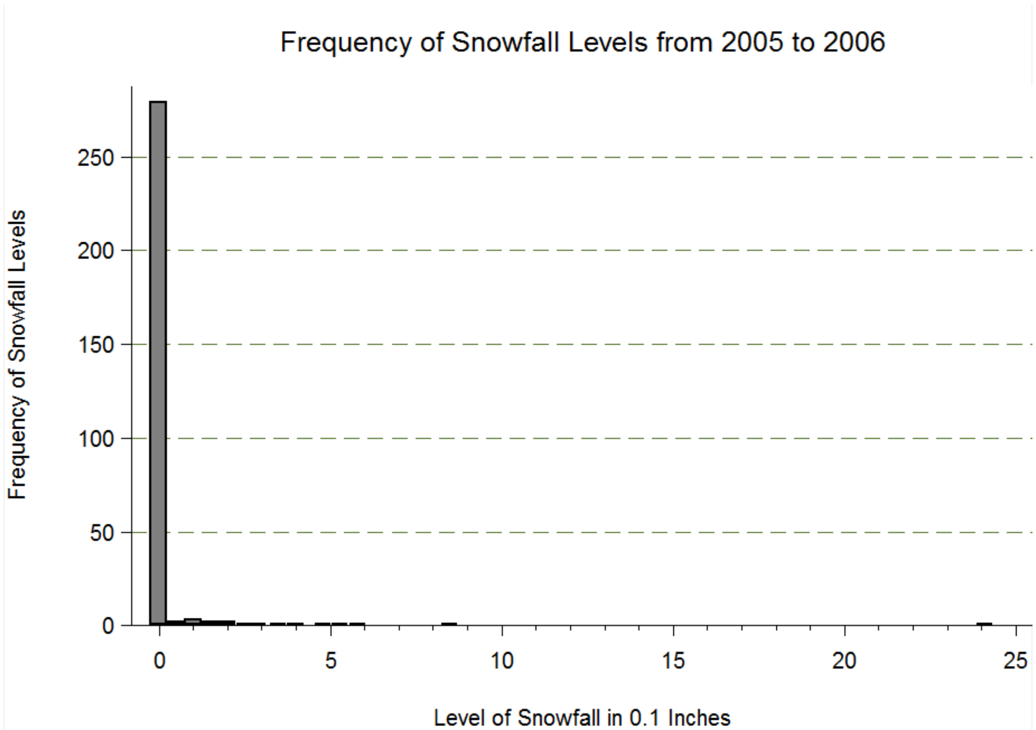
It is interesting to note from Table 3 that while Central Park snows more frequently in January, February is more prone to above average snowfalls.

4.2 Zero-Inflated Snowfall Data

Historically, 90% of the days between November 1st and March 31st of each year have experienced no snowfall. This fact renders typical quantitative methods useless,[19] where modelling strategies such as analyzing autocorrelation functions is not available due to the vast number of zeroes. For instance, the figure below illustrates the frequency of snow at different levels during the 2005 to 2006 winter season:

[19] Other models we considered include Intermittent Demand Analysis, Sparse Data Analysis, and related zero-inflated time series models used in biostatistics to model disease outbreak.

Figure 3



Please refer to the Appendix, Figures 1a to 1i for annual snowfall distributions from 2003 to 2013.

4.3 Treatment as Count Data

While the Roll orange juice paper was a major inspiration to this essay, a noteworthy impediment to our simply replicating his methodologies is that the nature of our datasets is different. Roll was predicting temperature, a value that can fluctuate up and down. Contrastingly, our futures contracts are based on snowfall *accumulation*, a value that can never decrease. We will employ lags of variables from futures contracts as regressors like Roll, but we will not be able to run OLS regressions.

Taking the above into account, we chose to model snowfall as count data. The main justification for treating snowfall as count data is as follows: our primary focus is not concerned with 'which day' in which snowfall occurred or 'how much snow fell during this particular month.' Instead, our main question is: How many inches of snowfall accumulated between November 1st and April 30th? As one will notice, the question was purposely devised so as to be answerable by our futures contract data.

This allows us to circumvent the use of zero-inflated time series techniques; furthermore, a count model has the added feature of taking into account the fact that snowfall cannot be negative.

Although the amount of snowfall is continuous, weather forecasts themselves typically report accuracy to a tenth of an inch. For instance, actual snowfall of 20.53 inches would naturally be rounded to 20.5 inches in published meteorological reports.

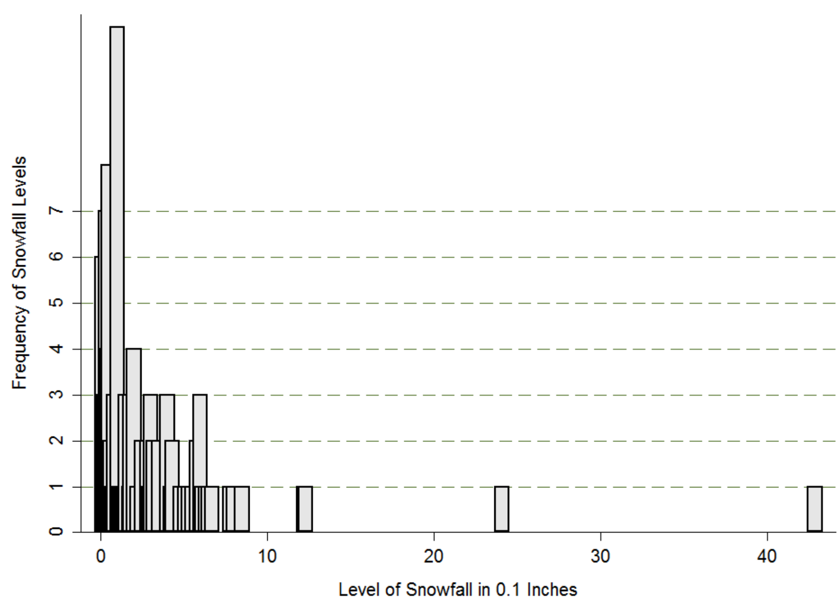
Keeping the above goal in mind, we propose to render the data as follows: each 0.1 inch of snowfall will be deemed as one 'count.' For instance, a day reporting 2.1 inches of snowfall will have experienced 0.1 inches of snowfall 21 times. Similarly, a day reporting 30.1 inches of snowfall will likewise be interpreted to have counted 0.1 inches of snowfall 301 times. Since we are not particularly interested in the rate of snow falling over certain hours in a day, nor do we have data on snowfall patterns of high and low during a *particular* day (eg. we do not have any descriptive evidence that it snows more during the night, merely that it *snowed* a certain amount), we believe our choice of rendition to be a reasonable simplification that will allow us to model snowfall. Additionally, using a small unit of 0.1 inches rather than 1 or 2 inches, provides enough granularity to model snowfall levels as given by weather reports.

One caveat is that simplifying snowfall to discrete data may potentially round off days where snowfall is less than or equal to 0.09 inches to 0. In other words, our count data may significantly misrepresent the number of days that snowed; however upon examining historical daily snowfall from 1940 to the present (again, between the days of November 1 and March 31 of each year), we found that of the days that *did* snow there was never a record between 0 and 0.09 inches. We will use this ample historical evidence going forward and assume that the lowest amount of snowfall on any given day is 0.1 inches.

4.4 Graphing Snowfall Data: A Fishy Distribution?

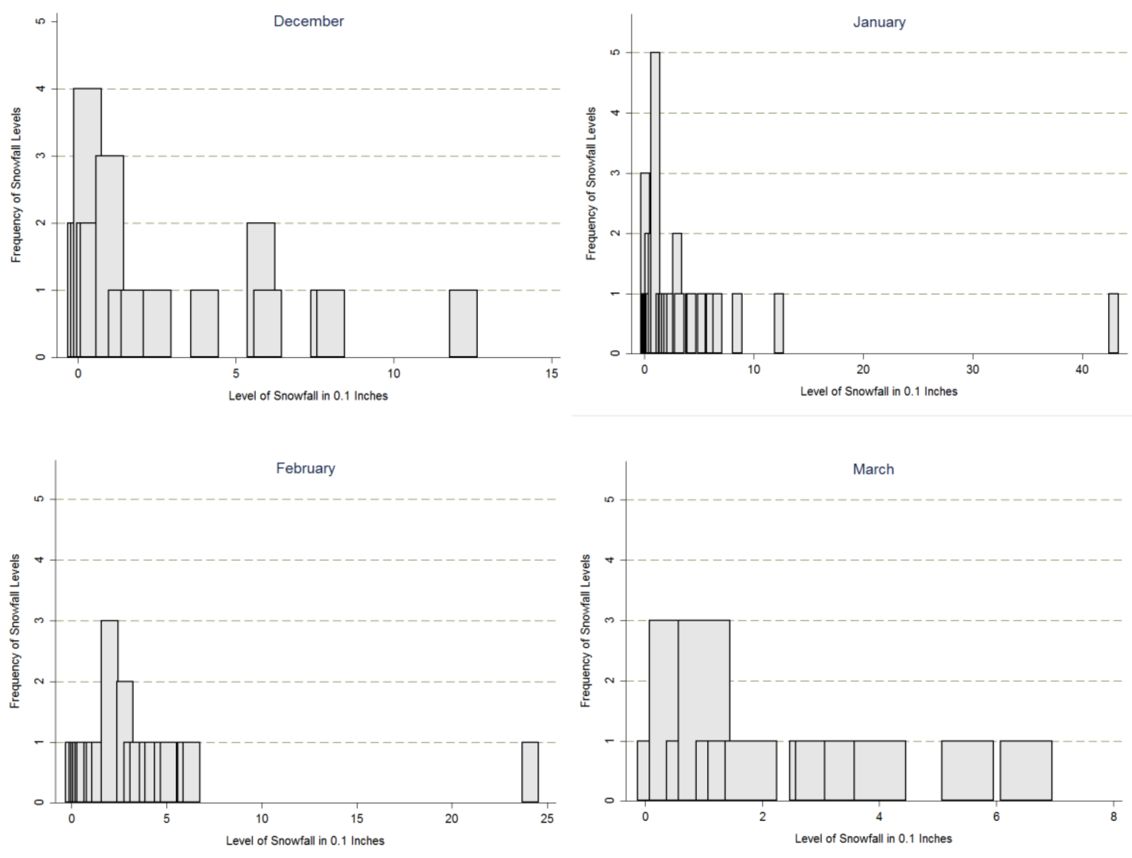
In order to model snowfall count data, we first attempted to model the relationship between snowfall and different contracts' movements by appealing to a Poisson regression. The main rationale to starting with a Poisson model is that merely *plotting* snowfall levels, conditioned on snowfall occurring, reveals a Poisson-esque, positively-skewed distribution.

Figure 4 *Conditional Distribution of Snowfall Frequency, Years 2003 to 2013*



If we examine the data more granularly on a monthly basis, we see the same indicators of a Poisson distribution, albeit more weakly:

Figure 5a to 5d *Conditional Distribution of Snowfall Frequency by Months, 2003 to 2013*



From these visual representations, it seems advisable for us to begin with a Poisson regression.

4.5 Prediction Variables

Aside from identifying a possible distribution, we require regressors that we suspect to be the best for predictions. For a successful net result, we must catch the most *informative* regressors. Referring back to the figures below as an example, the futures' behaviour during the 2005 to 2006 season provides some insight into our approach:

From Figure 1, it is evident that the contract which pays off if snow accumulates anywhere between 40.0 inches and 49.9 inches was considered the clear winner well before the end of the season; furthermore, it should be highlighted that it had been considered the winner since February 12th, which is 78 days in advance of contract termination. There also exist notable fluctuations from the 20+ and 30+ inches contracts. The termination of the 30+ inches contract on February 12th corresponds to a 24.1 inches snowfall which occurred on that same day, bringing up the total snowfall accumulation count to 38 inches.

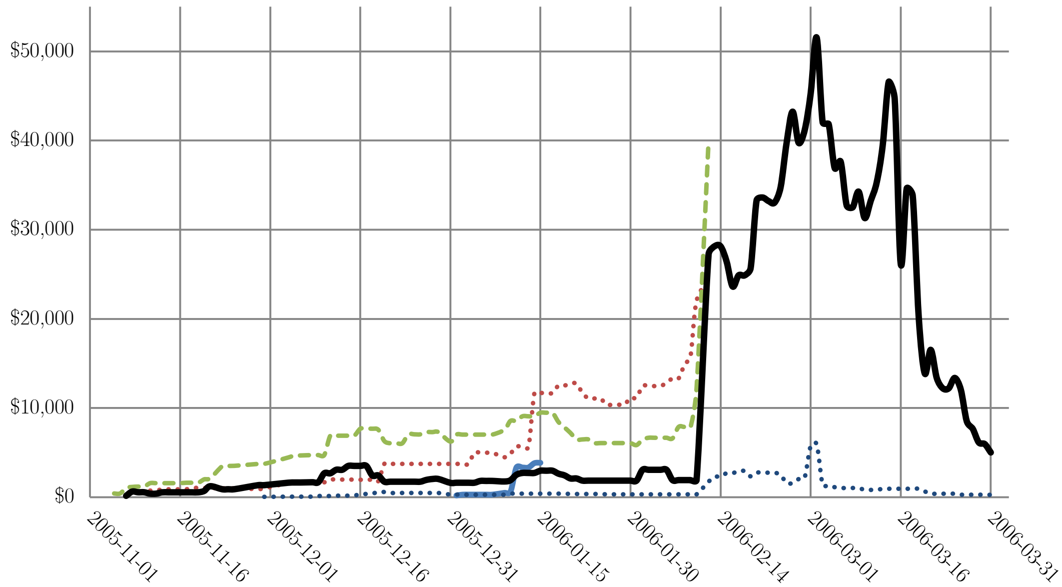
The above description supports what should already be intuitive: that the contracts' variables move in tandem with snowfall fluctuations. The question is this: which characteristics of the contracts provide the most information?

The following variables were used in our regressions because they concluded a clear winner when graphed: Closing Price, Open Interest, and Value. Value was created because we discovered that in the several instances where we added Value to a model composed of solely Closing Prices and Open Interest, the addition significantly improved the fit of the model. This improvement was identified by running LR tests.

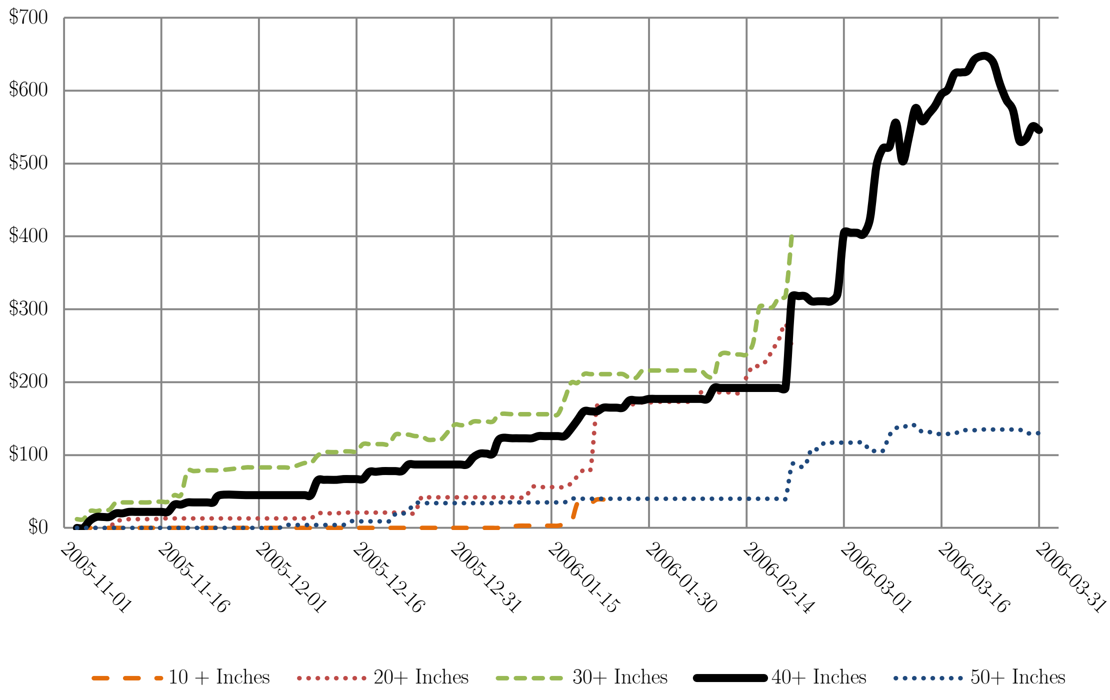
In the following Figures 6a and 6b, we see the 40+ Inches contract rise above the other contracts, indicating that it is believed to be most likely to occur. Please see Appendix, Figures 2a to 4c for graphs of contract' Closing Price, Open Interest, and Value for the sample of years 2003 to 2008.

Figure 6a to 6c

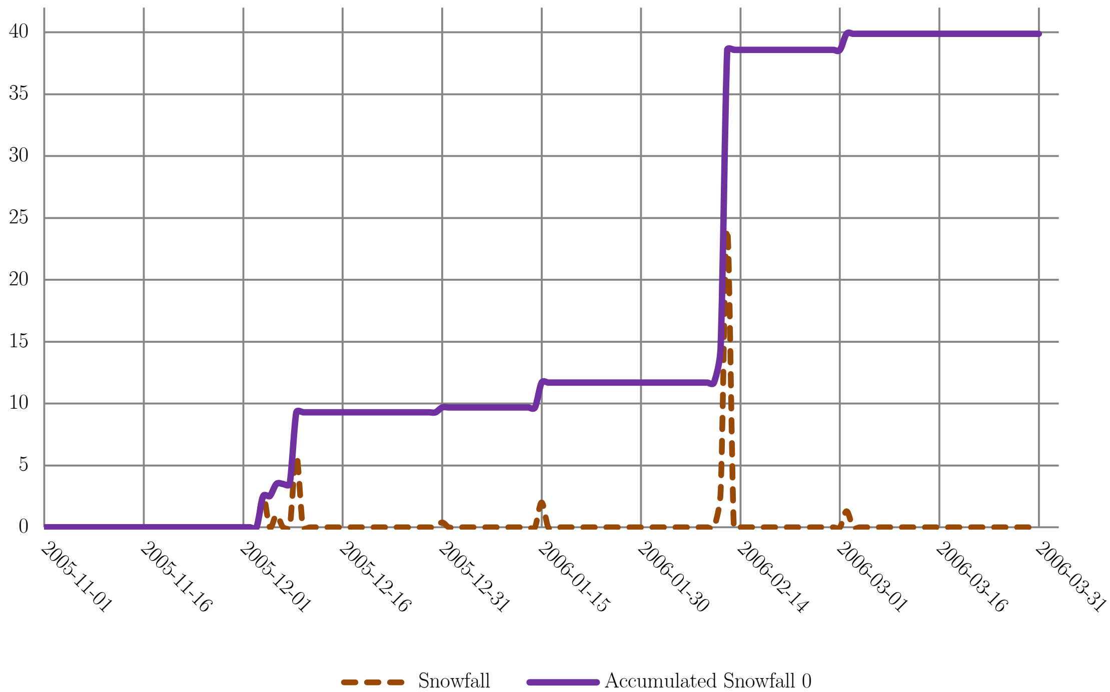
Futures Contract Values, 2005-06



Futures Open Interest, 2005 - 06



Snowfall and Accumulated Snowfall, 2005 - 06



Contrastingly, the variables Session High, Session Low, Lifetime High, Lifetime Low Session Open, and Trading Volume, when graphed, were identified as poor indicators for determining the winning contract. At any point in time, too many fluctuations occur between the different contracts, and thus do not clearly indicate a winner (eg. if a contract's session open was constantly fluctuating to zero, then this does not tell us anything, just that is it volatile). In particular, Trading Volume makes sense as a deficient proxy of information because it merely illustrates the amount of contracts that actually transact each day. The implication, of course, is that it does not take into account contracts that were *unable* to find counterparty.

Please see Appendix, Figures 5a to 6b, for sample graphic representation of the variables Session High and Trading Volume.

4.6 Poisson Regression Estimation

Models at different lags using different types of contracts were tried.[20] Since we do not have a particular hypothesis regarding which contracts would be the best predictor of future snowfall, our general approach was to run a 'catch-all' regression first and then remove insignificant variables.

However, before running any regressions, the 60+ inches contract was excluded from all years because from our graphs, we see that it almost never varies and therefore conveys very little information about what traders think about snowfall. The reason we tried the regression at different lags is because later, we will attempt to determine how far into the future for which closing prices have predictive powers.

Model 1: Closing Price[21]

	No Lag	Lag 1	Lag 2	Lag 3	Lag 4
close10	0.0122	0.0049	0.0062	0.0094	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
close20	0.0053	-0.0199*	-0.0067	-0.005	-0.0086
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
close30	0.0164	0.0435	0.0021	0.0049	0.0274
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
close40	-0.1254	-0.1264*	-0.1348**	-0.1254*	-0.1268*
	(0.12)	(0.06)	(0.04)	(0.05)	(0.05)
close50	0.0771	0.0862	0.0956	0.0921	0.0813
	(0.09)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	0.0937	1.0231	1.2655**	0.8737	0.7466
	(0.83)	(0.58)	(0.44)	(0.58)	(0.69)
Obs.	56	56	57	57	57
chi2	5.39	23.02	15.66	13.35	16.43
BIC	387.46	338.88	328.82	347.47	346.61
Pearson	482.9248	331.6693	226.8973	266.4998	259.9568

[20] In other words, types of models being composed solely of either contracts' Closing Prices, Open Interests, or Values.
[21] Standard errors are reported in brackets. Coefficients with significance to 0.05 denoted with *, significance to 0.01 denoted with **, and significance to 0.001 denoted with ***.

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
close10	0.0075	-0.0057	-0.0037	0.0043	-0.0123
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
close20	-0.0085	-0.0199	-0.0086	0.0057	-0.0017
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
close30	0.0244	0.0589	0.0597	0.0043	0.0536**
	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
close40	-0.1343**	-0.1080**	-0.0831**	-0.0901***	-0.0703***
	(0.05)	(0.04)	(0.03)	(0.02)	(0.01)
close50	0.0882	0.0692	0.0224	0.0874***	0.0443**
	(0.05)	(0.04)	(0.02)	(0.02)	(0.02)
Constant	0.8143	1.3410*	0.8453	0.373	0.7917
	(0.59)	(0.68)	(0.54)	(0.35)	(0.47)
Obs.	57	60	61	60	61
chi2	18.18	15.39	21.37	98.02	49.02
BIC	339.75	353.77	339.43	280.83	19.19
Pearson	245.0128	261.2661	239.8488	143.3224	180.7143

Model 2: Open Interest

	No Lag	Lag 1	Lag 2	Lag 3	Lag 4
open10	0.0059	0.0066	0.0032	-0.0002	0.0015
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
open20	0.0033	0.0026	0.0052	0.0118***	0.0123***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
open30	-0.0218*	-0.0222*	-0.0198*	-0.0207*	-0.0218**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open40	0.0269*	0.0293*	0.0270*	0.0303*	0.0322*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open50	-0.0064	-0.0101	-0.0099	-0.0164	-0.0203*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.9011**	0.8993**	0.8655**	0.5609	0.5274
	(0.30)	(0.30)	(0.33)	(0.32)	(0.30)
Obs.	58	59	60	60	60
chi2	96.22	50.69	28.4	52.27	44.29
BIC	333.84	339.47	359.34	338.73	333.97
Pearson	196.6828	198.025	224.7223	196.4744	187.4663

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
open10	-0.005	-0.0095	-0.0071	-0.0088	-0.0084
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
open20	0.0134***	0.0065**	0.0078**	0.0065***	0.0083***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
open30	-0.0210**	-0.0113	-0.0198*	-0.0141	-0.0211*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open40	0.0340*	0.0184	0.0315	0.0096	0.0400*
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
open50	-0.0143	-0.0017	-0.0151	0.0069	-0.0261
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Constant	0.5242	1.0244**	1.0667***	1.3307***	0.9980***
	(0.29)	(0.33)	(0.23)	(0.38)	(0.20)
Obs.	60	63	65	66	67
chi2	83.55	21.83	10.42	19.19	14.39
BIC	320.33	406.47	390.43	430.57	391.16
Pearson	173.7358	314.9056	251.0031	347.4961	244.875

Model 3: Value

	No Lag	Lag 1	Lag 2	Lag 3	Lag 4
value10	0.0001	0.0001	0	0	0.0001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value20	0	0	0.0001	0.0001***	0.0002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value30	-0.0003	-0.0003	-0.0002	-0.0003*	-0.0004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value40	0.0003	0.0004	0.0003	0.0004*	0.0006**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value50	-0.0001	-0.0001	-0.0001	-0.0003*	-0.0004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.9433**	0.8729**	1.0449***	0.5920*	0.4088
	(0.34)	(0.32)	(0.32)	(0.28)	(0.29)
Obs.	58	59	60	60	60
chi2	92.21	42.69	31.58	42.27	31.74
BIC	354.35	362.43	375.86	344.53	324.71
Pearson		247.0029	268.4271	214.8568	195.7805

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
value10	-0.0001	-0.0001	0	0	-0.0001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value20	0.0002***	0.0001	0	0	0.0001**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value30	-0.0003*	-0.0001	-0.0002*	-0.0001	-0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value40	0.0005*	0.0002	0.0002	0.0002	0.0003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
value50	-0.0002	-0.0002	0	-0.0002	-0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.5735*	1.2048	1.3968***	1.3669*	1.1860***
	(0.27)	(0.75)	(0.36)	(0.64)	(0.31)
Obs.	60	45	65	49	67
chi2	75.65	8.75	8.89	7.73	15.31
BIC	328.91	369.34	441.23	391.66	441.64
Pearson	195.3574	401.0949	411.0352	411.1509	407.1374

At this point, it is advisable to stop running the Poisson regressions. For *every* Poisson regression that we have run so far, the Pearson goodness-of-fit test[22] indicates that the Poisson model is not a good fit for our data.

4.7 The Problem of Overdispersion

It became apparent that overdispersion was present, implying that our Poisson model does not fit well. Poisson models implicitly assume that the conditional mean of the dependent variable and its conditional variance are equal. As Cameron and Trivedi (2009) note, the fundamental limitation of the Poisson model is that it characterizes its entire distribution in terms of one scalar parameter, the mean (μ). If a count model is ill-fitted by the Poisson, an improvement may be brought about by seeking a model containing more information – or said another way, one with more *parameters*. Failure to correct for overdispersion results in standard error estimates that are too small and results in too much confidence in the results.

[22] The Pearson goodness-of-fit chi-squared test calculates a test static and a p-value for was significant for *every* Poisson model we ran. In other words, [Prob. > Chi2(df)] = 0.00 for every Poisson regression. STATA documentation suggests trying the Negative Binomial regression if this occurs.

For each year in our analysis, it is evident that the variance was always larger than the mean - a clear sign of overdispersion.

Table 4 *An Illustration of Overdispersion in the Data*

	2005-06	*2006-07*	*2007-08*	*2008-09*	*2010-11*	*2011-12*
Mean	3.29	1.02	0.97	2.28	5.12	0.37
Variance	518.53	30.56	6.18	78.45	379.70	15.30

In practice, an overdispersed Poisson model can be replaced with a Negative Binomial, which is known to be a more accurate model in cases of discrete, overdispersed data. In fact, the Negative Binomial distribution can be derived as a Poisson random variable.[23]In actuarial practises, it has been used to model the rate at which people encounter accidents based on weather.

The negative binomial model is a form of generalized linear model (GLM) for count data. It is important to use a GLM model because it allows for dependent variables (or response variable) with a non-normal distribution of errors. Intuitively, we would expect this to be true for snowfall, or any natural, weather-related phenomenon. We will assume that the count variable Y_i , the number of 0.1 inches of snow in a day, follows a Poisson-like process, but that the variation is greater than a true Poisson.

The first two moments of the Negative Binomial Model are as follows:

$$E(Y_i \mid \mu_i, \alpha) = \mu_i$$

$$VAR(Y_i \mid \mu_i, \alpha) = \mu_i + \frac{{\mu_i}^2}{\omega} = \mu_i \left(1 + \frac{\mu_i}{\omega}\right)$$

There is a constant scale parameter, ω. Unless ω is very large, the variance of Y_i in the negative binomial will increase faster than that of the Poisson. ω can also be interpreted

[23] It should be noted that we did not merely default to the negative binomial model after failing with the Poisson because using the Negative Binomial yields the same conditional mean.

as a sort of inverse of the amount of dispersion, α. The larger it gets, the more closely the model becomes a Poisson (where the variance of the count equals the mean). We are able to estimate the regression coefficients and parameter ω by way of maximum likelihood.

The expected count ${\mu_i}^*$ is the expected number of 0.1 inches of snow conditioned on it being a day i *with snow*. ${\mu_i}^*$ is a gamma-distributed, unobservable random variable with a mean of μ_i and constant scale parameter. For any observation i,

$${\mu_i}^* \sim \Gamma\left(\frac{1}{\alpha}, \alpha\mu_i\right) \sim \Gamma\left(\omega, \frac{\mu_i}{\omega}\right)$$

$$\text{where } \alpha > 0, \ \ \mu_i > 0$$

Therefore,

$$\begin{aligned} VAR(Y_i) &= E[VAR(Y_i|{\mu_i}^*)] + VAR[E(Y_i|{\mu_i}^*)] \\ &= E({\mu_i}^*) + VAR({\mu_i}^*) \\ &= \mu_i \left(1 + \frac{\mu_i}{\omega}\right) \end{aligned}$$

Additionally, as noted earlier, the negative binomial distribution has been used to model tornadoes. In chapter three of Simmons and Sutter's book *Economic and Societal Impacts of Tornadoes*, negative binomial regressions were run in order to analyse tornado casualties. They regressed the *number* of fatalities (yes, it is a count model) caused by tornadoes over $1950 - 2007$ on independent variables such as the intensity measure of a tornado, time of day for when the tornado occurred, and the month of the occurrence, among many others. The above examples indicate that it is not entirely far-fetched for us to model snowfall using a count data model.

The zero-inflated negative binomial model was also considered. Although quantitatively it may yield a better specification, we chose not to pursue it because the nature of our data suggests against the existence of any inflation factor.[24] [25] We thereby proceed with the Negative Binomial model because it is the best model available to us.

4.8 The Negative Binomial Model

We will be using a Negative Binomial Distribution[26]. Our model is as follows: Imagine that during the course of a day, an x-number of repeated trials occur. A trial can be held once every half-hour, once every five minutes, or any other hypothetical increments that matches the pace of change in weather conditions. The same Negative Binomial distribution exists for each day. The more snowfall we count on a given day, the more snowfall is accumulated over months towards the end result that determines the winning contract.

Assumption #1: Each trial can result in two possible outcomes:

Success with probability P: No Snow

Failure with probability $1 - P$: 0.1 Inches of Snow

Assumption #2: The probability of success, P, is the same on every trial.

Assumption #3: The trials are independent.

Assumption #4: The same number of trials occurs each day.

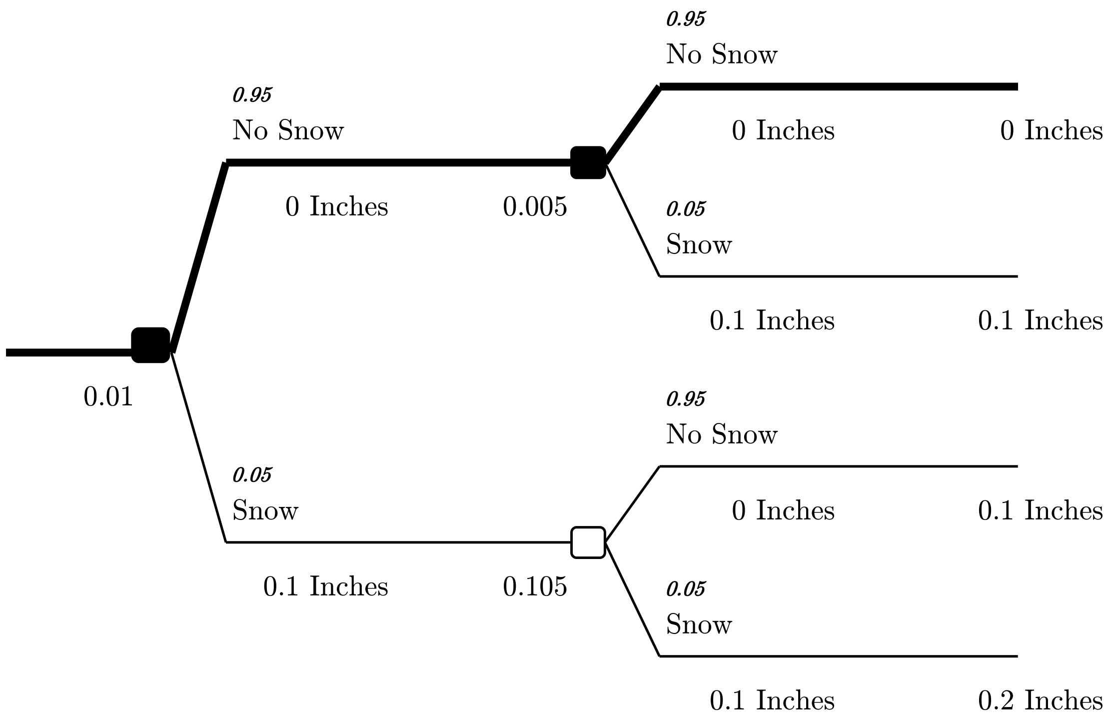
Taken together, the Assumptions 1 to 4 can be represented by the tree diagram below:

[24] In a zero-inflated negative binomial model, it is assumed that observed values are caused by two separate distributions: first, a binary process (inflation factor) of 0 or 1, and then a second, typical count process such as the Poisson or Negative Binomial. If the binary process yields 0, then the observed value is simply 0. If the binary process is 1, then the observed value will be from the count mode: 0, 1, 2, etc. Therefore, an observation of 0 may be the result of the binary process yielding 0, or due to the result of the count model yielding 0, conditioned on the binary process yielding 1.

[25] Spurious relations are a serious crime!

[26] Other commonly-used names include the Pascal, or Polya distribution.

Figure 5 *Tree Diagram of Snowfall in One Day, if P = 0.95*



The tree diagram maps a subset of nodes that can occur in a day. As illustrated, only one branch will finish with no snow at all during the entire day. All other branches, aside from the branch yielding no snow, together represent a distribution of snowfall, conditioned on snow falling at all. This simple model accurately captures the fact that zero snow is the most likely outcome, and that smaller amounts of snowfall is more likely than larger amounts.

Out of the above four assumptions, Assumption 2 is probably the most questionable. If it is satisfied, this would imply that subsequent snowfalls after a snowfall are equally likely as snowfalls occurring after a period of no snow. This may conflict with intuition since one could argue that if snow actually occurred, then the weather conditions must have made it more likely for it to snow in that moment and in any subsequent moment in time close to it.

While a valid argument, we would like to point out the simple observation that historically, only 10% of the days 7,260 days we surveyed that occurred between November and March since 1940 have snowed. It is clearly more likely for it to *not* snow on any given day, despite cycling through numerous trials, meaning that probability P (probably of not snowing) must be greater than probability $1 - P$. Even if it were to snow in a trial, we will assume that it only makes it slightly more likely for it to snow in the next trial, insignificantly lowering the large probability P slightly.

5. Regression Results

5.1 Negative Binomial Regression Results

The Negative Binomial regression utilizes the method of maximum likelihood in order to estimate coefficients. Within all the regression tables below, the parameter Alpha – or dispersion parameter - is always greater than 0, confirming our suspicions of data overdispersion and validating our use of the Negative Binomial model.[27]

In order to determine the significance of model, the p-values of a Wald chi-square statistic was used. It tests whether all coefficients are simultaneously equal to zero. Additionally, in order to compare the fit of different specifications that passed the Wald test, we relied mainly on the Bayesian Information Criterion (BIC), which presents the key advantages of firstly, controlling for overfitting and secondly, having the ability to compare models that are not nested. The BIC allows us to directly compare specifications with one another, where a smaller BIC statistic indicates that one particular model is better relative to another.[28] A composition of futures prices from various lags could all be

[27] If Alpha = 0, then the model reduces to a Poisson model. The fact that it is > 0 indicates that using a Poisson is incorrect.

[28] When comparing the BIC statistics for two models, the interpretation of the magnitude of absolute difference is as follows: 0 – 2 (Weak), 2 – 6 (Positive), 6 – 10 (Strong), >10 (Very Strong).

potential candidates as regressors within the same regression depending on the level of information perforation reflected in the BIC.

5.2 Model 1: Closing Price

Similar to the process with which we chose models to regress with the Poisson model, we will do the same with the Negative Binomial model:

	No Lag	Lag 1	Lag 2	Lag 3	Lag 4
close10	0.0134	0.0107	0.0091	0.0098	0.0104
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
close20	0.0017	-0.0036	-0.002	-0.0026	-0.0043
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
close30	0.018	0.0192	-0.0115	0.002	0.0227
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
close40	-0.0636	-0.0733	-0.1056**	-0.0893*	-0.0935*
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
close50	0.0242	0.0512	0.0753*	0.0664	0.0571
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Constant	-0.1421	0.1951	1.0260**	0.6546	0.3408
	(0.88)	(0.56)	(0.32)	(0.45)	(0.54)
Obs.	56	56	57	57	57
chi2	7.5	13.44	11.13	8.91	10.77
alpha	1.13	0.98	0.72	0.86	0.86
BIC	265.34	259.8	255.23	261.29	260.74
Pearson	114.27	84.01	54.55	66.47	64.12

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
close10	0.0109	-0.0007	-0.0002	0.0108	-0.0017
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
close20	-0.0025	-0.004	0.0025	0.0082	-0.0055
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
close30	0.0143	0.0293	0.039	0.0036	0.0469*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
close40	-0.1089**	-0.0737**	-0.0751***	-0.0854***	-0.0591***
	(0.04)	(0.02)	(0.02)	(0.02)	(0.01)
close50	0.0679	0.0473	0.0226	0.0683***	0.0256
	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
Constant	0.4328	0.7907	0.4575	-0.1009	0.4164
	(0.47)	(0.65)	(0.52)	(0.37)	(0.55)
Obs.	57	60	61	60	61

chi2	10.33	11.67	15.14	36.11	27.64
Alpha	0.81	0.9	0.69	0.53	0.68
BIC	258.52	273.97	270.53	257.51	269.17
Pearson	59.42	65.95	60.62	41.34	51.88

By the BIC and Pearson statistic, Closing Prices lagged at three days seem to be the best predictor of snowfall. It is also noteworthy to point out that the 40+ Inches contract was significant as a regressor for regressions starting from Lag 2 and beyond.

5.4 Model 2: Open Interest

	No Lag	**Lag 1**	**Lag 2**	**Lag 3**	**Lag 4**
open10	0.001	0.0029	-0.0013	-0.0016	-0.0007
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
open20	0.0054*	0.0043*	0.0087***	0.0126***	0.0128***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
open30	-0.0171**	-0.0177**	-0.0169**	-0.0188**	-0.0193**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open40	0.0209*	0.0231*	0.0238**	0.0272**	0.0284**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open50	-0.002	-0.0063	-0.0089	-0.0149	-0.0177
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.8085*	0.8137*	0.6785*	0.5171	0.5143
	(0.33)	(0.34)	(0.32)	(0.31)	(0.31)
Obs.	58	59	60	60	60
chi2	67.83	45.39	18.26	60.44	52.86
Alpha	0.75	0.73	0.75	0.64	0.63
BIC	259.64	265.92	271.9	267.42	267.07
Pearson	46.57	47.93	47.98	46.76	44.56

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
open10	-0.0058	-0.0129*	-0.0053	-0.002	-0.0071
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
open20	0.0140***	0.0092**	0.0076**	0.0098**	0.0081**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
open30	-0.0186**	-0.0093	-0.0173**	-0.0175*	-0.0154*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open40	0.0297**	0.0179	0.0281*	0.0162	0.0290**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
open50	-0.0137	0.001	-0.0152	-0.0023	-0.0208*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.5054	0.7794**	0.9910***	1.0132***	0.9621***
	(0.31)	(0.29)	(0.25)	(0.27)	(0.22)
Obs.	60	63	65	66	67
chi2	69.37	24.22	15.63	14.91	18.72
Alpha	0.56	0.88	0.79	0.89	0.75
BIC	262.73	288.41	295.87	306.45	306.69
Pearson	42.58	62.83	54.96	63.69	58.10

We make note that for the majority of the regressions, Open Interest for the 20+, 30+, and 40+ Inches contracts were significant regressors.

5.4 Model 3: Value

	No Lag	Lag 1	Lag 2	Lag 3	Lag 4
10+ inches	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
20+ inches	0.0001*	0	0.0001**	0.0001***	0.0002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
30+ inches	-0.0002*	-0.0002*	-0.0002**	-0.0002**	-0.0003**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
40+ inches	0.0002*	0.0003*	0.0002	0.0003*	0.0004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
50+ inches	0	-0.0001	0	-0.0002	-0.0003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.8342*	0.7451*	0.9051**	0.6040*	0.5015
	(0.35)	(0.34)	(0.28)	(0.31)	(0.32)
Obs.	58	59	60	60	60
chi2	61.25	35.21	17.12	54.03	49.34
alpha	0.84	0.82	0.81	0.7	0.67
BIC	263.19	269.67	273.88	270.29	269.61
Pearson	56.76	59.7	53.65	52.70	52.23

	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
10+ inches	-0.0001	-0.0001	0	-0.0001	-0.0001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
20+ inches	0.0002***	0.0001*	0.0001	0.0001	0.0001**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
30+ inches	-0.0002**	-0.0001	-0.0002*	-0.0001	-0.0002*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
40+ inches	0.0003*	0.0004	0.0002	0.0001	0.0003*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
50+ inches	-0.0001	-0.0003	-0.0001	-0.0002	-0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.6455*	1.0216*	1.3297***	1.2830*	1.1252***
	(0.29)	(0.52)	(0.32)	(0.57)	(0.30)
Obs.	60	45	65	49	67
chi2	55.32	9.73	11.13	4.27	20.48
Alpha	0.64	1.14	0.98	1.1	0.91
BIC	266.99	229.52	305.03	252.78	314.26
Pearson	49.63	70.98	84.12	88.16	92.92

For Value, it appears that the 20+ and 30+ Inches Contracts are significant at lags of five days or less.

In addition to the above three models, we also ran what we call a 'Combined' model composed of all significant regressors from the previous three models.

5.5 Model 4: Combination of Significant Regressors from Models 1 to 3

	No Lag		Lag 1		Lag 2
open 20+	-0.0522	open 20+	-0.0071	close 40+	-0.0491
	(0.04)		(0.01)		(0.03)
open 30+	-0.0833***	open 30+	-0.0243*	close 50+	0.012
	(0.02)		(0.01)		(0.03)
open 40+	0.1783***	open 40+	0.0220**	open 20+	0.0068
	(0.05)		(0.01)		(0.01)
value 20+	0.0006	value 20+	0.0001	open 30+	-0.0123
	(0.00)		(0.00)		(0.01)
value 30+	0.0008**	value 30+	0.0001	open 40+	0.0143*
	(0.00)		(0.00)		(0.01)
value 40+	-0.0018***	Constant	1.0599***	value 20+	0
	(0.00)		(0.28)		(0.00)
Constant	0.8643***			value 30+	0
	(0.22)				(0.00)
				Constant	1.2587***
					(0.32)
Obs.	58	*Obs.*	58	*Obs.*	57
chi2	62.44	*chi2*	31.69	*chi2*	21.45
Alpha	0.46	*Alpha*	0.74	*Alpha*	0.62
BIC	251.95	*BIC*	258.43	*BIC*	257.01
Pearson	34.36	*Pearson*	44.77	*Pearson*	37.48

	Lag 3		Lag 4		Lag 5
close 40+	-0.016	close 40+	-0.0061	close 40+	-0.0153
	(0.01)		(0.02)		(0.02)
open 20+	0.0161	open 20+	0.0142	open 20+	0.0377*
	(0.01)		(0.02)		(0.02)
open 30+	-0.0569	open 30+	-0.0539	open 30+	-0.0769**
	(0.03)		(0.04)		(0.03)
open 40+	0.087	open 40+	0.0787	open 40+	0.0891**
	(0.06)		(0.05)		(0.03)
value 20+	-0.0001	value 20+	-0.0001	value 20+	-0.0003
	(0.00)		(0.00)		(0.00)
value 30+	*0.0005*	value 30+	0.0005	value 30+	0.0008*
	(0.00)		(0.00)		(0.00)
value 40+	*-0.0009*	value 40+	-0.0008	value 40+	-0.0010**
	(0.00)		(0.00)		(0.00)
Constant	0.9027***	Constant	0.8619**	Constant	0.8181**
	(0.25)		(0.27)		(0.25)
Obs.	60	*Obs.*	60	*Obs.*	60
chi2	24.59	*chi2*	28.8	*chi2*	34.97
Alpha	0.61	*Alpha*	0.69	*Alpha*	0.54
BIC	274.79	*BIC*	276.73	*BIC*	270.4
Pearson	40.67	*Pearson*	38.51	*Pearson*	36.5

	Lag 10		Lag 15		Lag 20
close 40+	-0.0295	close 40+	-0.0199	close 40+	-0.0303*
	(0.02)		(0.02)		(0.01)
close 10+	-0.0074	open 20+	0.0037*	close 50+	0.0539**
	(0.01)		(0.00)		(0.02)
close 20+	0.0146	open 30+	-0.012	open 20+	0.0007
	(0.01)		(0.01)		(0.00)
close 20+	0	open 40+	0.0097*	open 30+	-0.0002
	(.)		(0.00)		(0.00)
Constant	1.1627	value 30+	0	Constant	0.8228***
	(0.91)		(0.00)		(0.19)
		Constant	1.4846***		
			(0.34)		
Obs.	63	*Obs.*	65	*Obs.*	85
chi2	3.34	*chi2*	13.48	*chi2*	11.64
Alpha	1.18	*Alpha*	0.84	*Alpha*	0.83
BIC	294.61	*BIC*	298.82	*BIC*	393.83
Pearson	94.46	*Pearson*	67.1	*Pearson*	99.82

	Lag 25
close 30+	0.0335**
	(0.01)
close 40+	-0.0356**
	(0.01)
open 20+	0.0338
	(0.02)
open 30+	-0.0462*
	(0.02)
open 40+	0.0412
	(0.02)
open 50+	-0.0202**
	(0.01)
value 20+	-0.0003
	(0.00)
value 30+	0.0004
	(0.00)
value 40+	-0.0002
	(0.00)
Constant	0.3478
	(0.39)
Obs.	67
chi2	63.36
Alpha	0.49
BIC	307.14
Pearson	43.94

Overall, the combination model yielded the best, if not better, model at each lag. Again, we appeal to the BIC and Pearson statistic in order to draw this conclusion. At this point, we have systematically run several preliminary regressions. From our results, we are able to determine a model within each lag that fares better than others.

6. Meteorological Forecasts

6.1 Accuracy

We were able to obtain professional, meteorological forecasts that forecasted the amount of snowfall to occur in the next twenty-four hours. Given that we have scientific forecasts, market forecasts, and the actual level of snowfall that occurred, we are able to compare how well our market forecast fared next to snowfall forecasts conducted by professional meteorologists.

First off, let us examine a subset of the data. Below lies a sample of snowfall and forecast data from the 2003 to 2004 winter season. As we can see, meteorological forecasts are mediocre at best.

Table 5 *Subset of Meteorological Forecasts from Dark Sky*

	Snowfall	Forecast		Snowfall	Forecast
December 5th, 2003	6.0	3.387	*January 14th, 2004*	0	0.531
December 6th, 2003	0	1.808	*January 15th, 2004*	0	2.058
December 13th, 2003	0	0.237	*January 16th, 2004*	0.2	0
December 25th, 2003	5.0	0	*January 17th, 2004*	0.5	0
January 5th, 2004	0.315	0	*January 26th, 2004*	0.9	0.138
January 6th, 2004	0.315	0	*January 27th, 2004*	0.9	4.810
January 8th, 2004	0.984	0	*January 28th, 2004*	0	4.020

All figures are denoted in inches.

The sample subset presented denotes the general accuracy observed from the data: it seems to be almost impossible to correctly forecast the exact inches of snow that will fall. It is pertinent that we reframe our question: What if we instead consider a successful meteorological forecast to be one where it correctly predicts whether *any* snow will occur at all? Below lies the frequency of this new, less stringent definition of successful forecast during the winter seasons of 2003 to 2008:

Table 6 *Frequency of Correct and Incorrect Meteorological Forecasts*

Year	2003/04	2004/05	2005/06	2006/07	2007/08
Correct Forecast	3	15	6	8	4
Incorrect Forecast	9	19	9	11	8

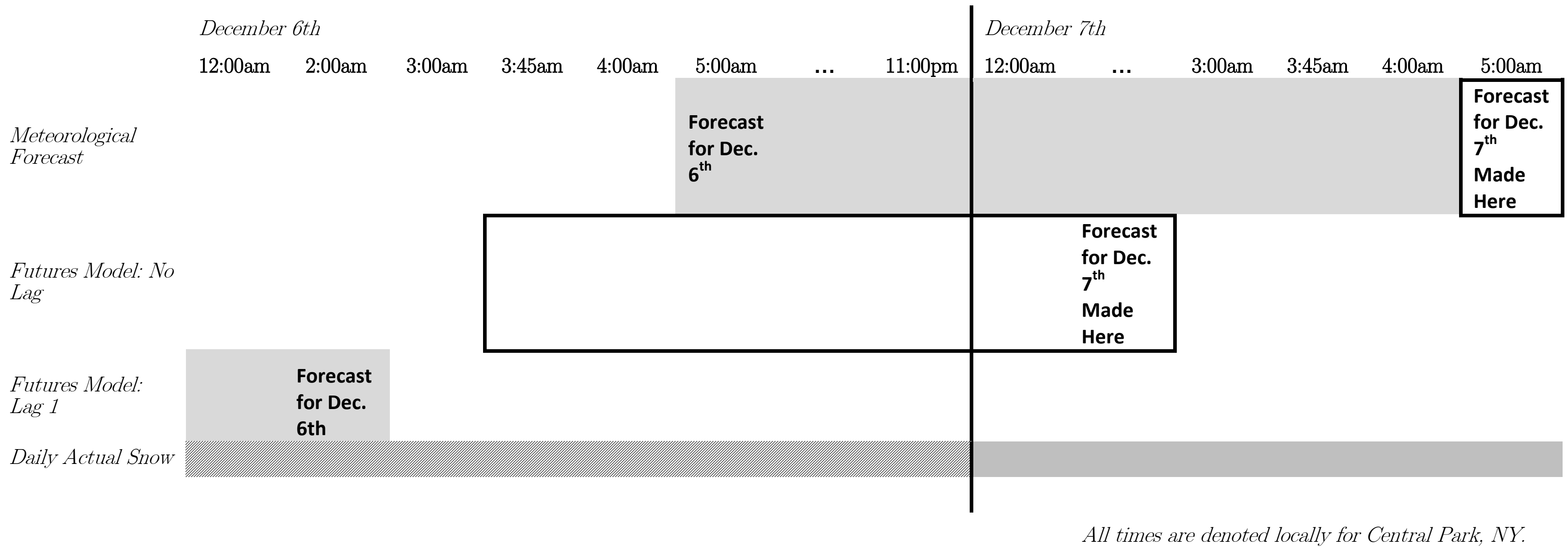
Evidently, professional forecasts have *some* ability in prediction, but it is mediocre at best. Knowing this, several hypotheses worth examining come to mind:

6.2 Hypothesis 1: Did the meteorologists or the futures market do a better job of forecasting?

Before attempting to answer this question, we must define a comparable futures contract to the meteorological forecast.

The earliest publicly-disseminated meteorological forecast for snowfall on December 7th is published at 5:00 am New York Time on December 7th. On the other hand, the futures market starts trading at 3:45am in New York each day (8:45am Ireland time, where the Intrade markets are legally registered) and ends at 3:00am the next day; therefore, the most accurate futures forecast for accumulated snowfall *made* on December 7th had its last trading period from 3:45am of December 6th through to 3:00am of December 7th. In other words, the closest, comparable futures model to the 24-hours ahead meteorological forecast is the futures forecast of the same day in question, the No Lag model.

Figure 7 *Example of Temporal Concerns for Various Forecast Methods for December 7th*



To test Hypothesis 1, we will compare regression models (conditioned on snow occurring) using our previously derived 'best model' with a regression model composed of solely the meteorological forecast. Before we do this, we examine the futures models we have had so far in order to determine the 'best' model comparable to the meteorological forecast. In the following comparison of models with No Lags, the Combination model has the smallest BIC and Pearson statistic, indicating that it is the model with the best fit amongst the four.

Table 7 *Comparison of No Lag Models*

	Closing Price	*Open Interest*	*Value*	*Combination*
Obs.	56	58	58	58
chi2	7.5	67.83	61.25	62.44
Alpha	1.13	0.75	0.84	0.46
BIC	265.34	259.64	263.19	251.95
Pearson	114.27	46.57	56.76	34.36

Therefore, we will be comparing results from the following models, namely:

Table 8 *Comparison Models and their Regressors*

	Regressand	*Regressor(s)*
Futures No Lag Model	Snowfall	Variables at No Lags: ▪ Open Interest of the 20+ Inches Contract ▪ Open Interest of the 30+ Inches Contract ▪ Open Interest of the 40+ Inches Contract ▪ Value of the 20+ Inches Contract ▪ Value of the 30+ Inches Contract ▪ Value of the 40+ Inches Contract
Meteorological Model	Snowfall	Meteorological forecast of snowfall to accumulate over the next 24 hours

The regression results are as follows:

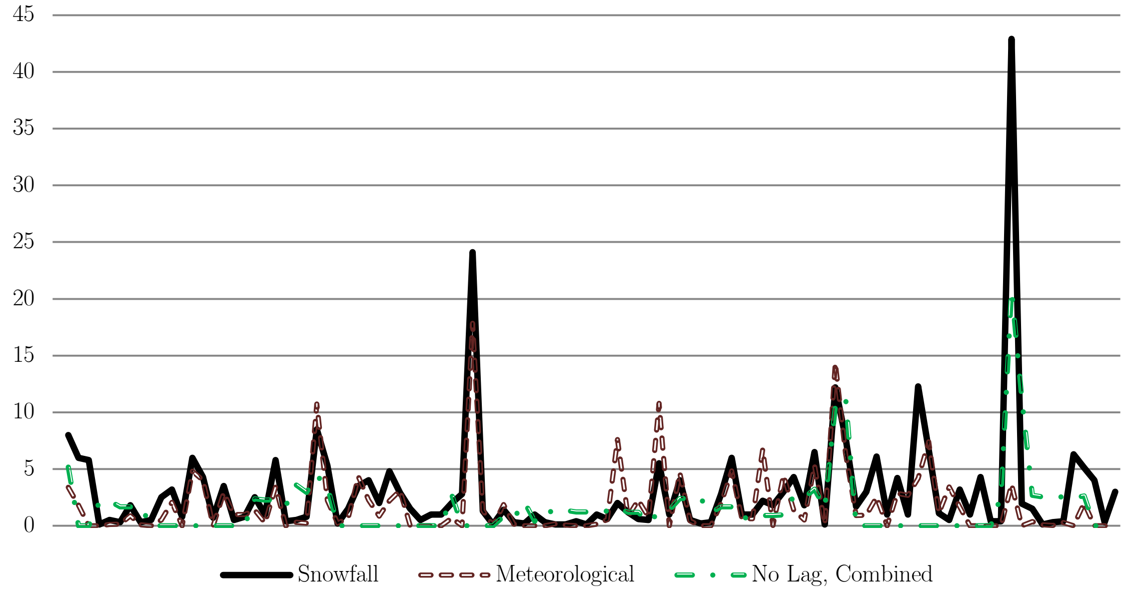
	Meteor Forecast		Futures Forecast
Meteor	0.2200***	Open 20+	-0.0522
	(0.06)		(0.04)
Constant	0.4727***	Open 30+	-0.0833***
	(0.13)		(0.02)
		Open 40+	0.1783***
			(0.05)
		Value 20+	0.0006
			(0.00)
		Value 30+	0.0008**
			(0.00)
		Value 40+	-0.0018***
			(0.00)
		Constant	0.8643***
			(0.22)
Obs.	99	*Obs.*	58
chi2	13.65	*chi2*	62.44
alpha	0.57	*alpha*	0.46
BIC	417.49	*BIC*	251.95
Pearson	133.84	*Pearson*	34.36

The regression results indicate that our futures model is a better fit. Specifically, the BIC and Pearson statistic are lower for the futures model.

Using the regression results, we now predict snowfall. We are interested to see which model generates values most closely matching actual snowfall. If traders' opinions are more accurate than the meteorological forecasts, then it would mean that there is significant private information being captured by the futures market. Contrastingly, if the meteorological forecasts are more accurate than the traders' opinions, then perhaps not all traders follow the weather forecast and there is information asymmetry present among participating traders.

Below, we graph the meteorological forecast and futures forecast on days that snowed.

Figure 9 *Actual Snowfall vs. Meteorological and Futures Forecasts (in Inches)*



Our forecast does not seem to fare too badly against the meteorological forecast. In order to obtain a precise comparison, the root mean squared error (RMSE) was calculated for each model.[29] The RMSE for the No Lag futures model was 3.77, while the RMSE for the Meteorological model was 5.47. Without a doubt, our futures forecast predicted more accurately than the meteorological forecast. This is an interesting result because the futures forecast is made two hours prior at 3:00am relative to the meteorological forecast at 5:00am.

In order to confirm our results, we conducted a simple examination of the predicted values from both models. On the snow days between December 2003 and March 31st, 2013, 94% of the meteorologically forecasted snowfall fell within one standard deviation of the snowfall sample standard deviation, conditioned on it being a snow day. On the other hand, 96% of the futures forecasts fell within one standard deviation of the snowfall sample standard deviation. Both are very close, and evidently, the predictive ability of the meteorological forecast should not be dismissed, which presents our motivation for Hypothesis 2:

6.3 Hypothesis 2: How much improvement is brought upon the futures model by including public information in our previous regression models?

This hypothesis is concerned with testing the existence of semi-strong form efficiency in futures markets. The meteorological forecast represents public information because it is available to all traders. If the addition of the term brings about a significant improvement in forecasting ability then we may take this as indication that not all traders took this information into account. In a sense, Hypothesis 2 deals with the efficiency of publicly available information.

---

[29] Root Mean Squared Error (RMSE) is defined to be $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(snowfall_i - prediction_i)^2}$. The RMSE was chosen for comparative purposes due to its greater sensitivity to the occasional large error, or deviation, from the actual snowfall.

In order to test this hypothesis, we ran our futures forecast model with the additional meteorological forecast as a new regressor. The results are as follows:

	Nested		Full
open 20+	-0.0522	open 20+	-0.0174
	(0.04)		(0.03)
open 30+	-0.0833***	open 30+	-0.0989***
	(0.02)		(0.02)
open 40+	0.1783***	open 40+	0.1433**
	(0.05)		(0.05)
value 20+	0.0006	value 20+	0.0002
	(0.00)		(0.00)
value 30+	0.0008**	value 30+	0.0010***
	(0.00)		(0.00)
value 40+	-0.0018***	value 40+	-0.0015**
	(0.00)		(0.00)
Constant	0.8643***	meteor	0.1957***
	(0.22)		(0.04)
		Constant	0.5690**
			(0.20)
Obs.	58	*Obs.*	58
chi2	62.44	*chi2*	103.81
alpha	0.46	*Alpha*	0.17
BIC	251.95	*BIC*	231.51
Pearson	34.36	*Pearson*	22.96

We ran the Wald Test to see whether the addition of meteorological forecasts significantly improves the model.[30] The p-value for the test was smaller than 0.000, indicating that the meteorological forecast does significantly improve the model.[31]

In fact, the additional meteorological forecast term made coefficients in the nested model more significant. After we predicted the values using the combined model, we found that

[30] We considered running the Likelihood Ratio (LR) test, but STATA did not allow us to do so given that our regressions were run with robust standard errors; however, when we ran the LR test with non-robust standard errors, we obtained the same conclusion as the Wald Test: that the Meteorological forecast variable improves our overall, forecast model. The test statistic was chi-square (df=1) = 24.50.

[31] The test statistic was chi-square(df=1) = 30.99.

the RMSE became 3.37 compared to the nested model (or entirely futures model) of 3.68.[32] BIC is now 231.51, and the Pearson statistic came to 22.96, both of which are improvements from the best-specified model we had from Section 1 (the Combination model, No Lags, which had BIC = 251.95 and Pearson = 34.36).

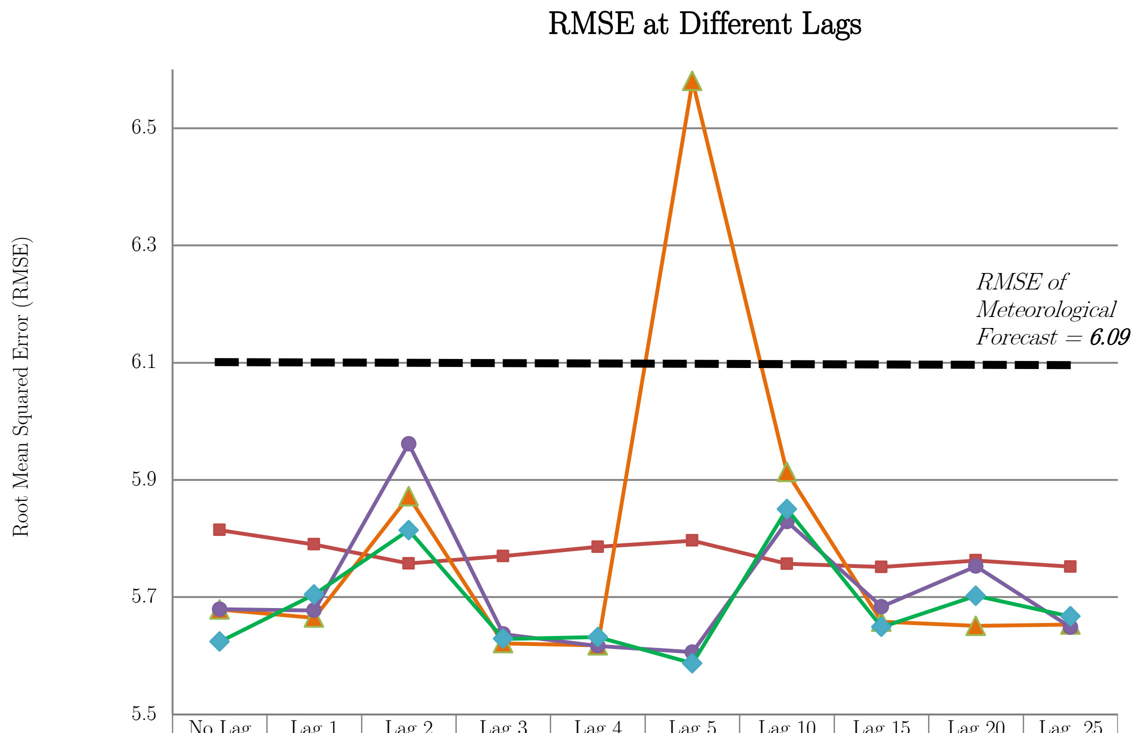
6.4 Hypothesis 3: How many lags does it take for the futures contract to predict on the same level as the meteorological contract?

It was determined that the 24-hours ahead meteorological forecast fared worse than the most comparable futures contract. This goes against intuition. It should be that the further into the future a prediction attempts to predict, the more likely it will be that the prediction is less accurate; however, as we saw from Hypothesis 1, the futures forecast made two hours before the meteorological forecast proved to be more accurate.

We now pose the following question: How many days ahead can we predict using futures contracts before the ability of the contracts become comparable to the prediction ability of the meteorological forecast? In other words: we would like to examine *at which futures lag* is the meteorological forecast comparable?

[32] The RMSEs for Hypothesis 2 were calculated with 58 observations instead of 61 observations. This was because some variables decrease in number of observations available due to the process of lagging it.

Figure 10 *Graph and Table of RMSE of Different Lags*



	No Lag	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 10	Lag 15	Lag 20	Lag 25
Close	5.81	5.79	5.76	5.77	5.79	5.80	5.76	5.75	5.76	5.75
Open	5.68	5.66	5.87	5.62	5.62	6.58	5.91	5.66	5.65	5.65
Value	5.68	5.68	5.96	5.64	5.62	5.61	5.83	5.68	5.75	5.65
Combined	5.62	5.70	5.81	5.63	5.63	5.59	5.85	5.65	5.70	5.67

Dashes represent the RMSE of the meteorological forecast.

According to the RMSE figures presented in Figure 10, futures contracts at *all* lags available to us are better predictors than the 24-hour ahead meteorological forecast. In other words, the futures market is sure to predict more accurately than the meteorological forecast up to 25 days in advance.[33] This result confirms the same result reflected in the BIC and Pearson statistic. In particular, the models composed of a

[33] It is important to note that we are calculating RMSE for days in which snowfall *occurred*. While this is consistent with our earlier methodologies and analysis of the conditional distribution, it also has the disadvantage of omitting error from the days where no snow fell, but where the meteorological forecast predicted positive snowfall. It, however, does not change the qualitative result of our regression,that our futures model predicts better than the meteorological model. Adding the days we are omitting will simply raise the RMSE of the meteorological model.

combination of Closing Prices, Open Interest, and Value seem to fare consistently well relative to other contract.

It is interesting to note that the model composed solely of the Closing Prices of contracts exhibited the peculiar overall trait of slightly decreasing RMSE as the number of lags increases. In other words, this model returns lower prediction errors the *further* into the future one forecasted. The RMSE of the Open Interest, Value, and the Combined model fluctuate up and down, but remain around the same mean overall. Please see Appendix, Figures 7a to 7d for the individual RMSEs of different models along with their trend lines. While the pattern of Closing Price's RMSEs may seem counterintuitive it may be that the nearer to a date, the noisier Closing Price becomes. More information is revealed day by day because with each day that passes, it becomes less and less likely for the accumulated snowfall to become more than the current level. Traders may be responding to this information. The examination of this hypothesis is beyond the scope of our discussion.

7. Conclusion

Information markets are a marvelous invention that does not receive enough attention. As we have shown, even phenomena that is supposed to be exogenous such as weather can be better predicted using information markets up to 24 hours better than the professional, meteorological forecast. Amongst this finding, perhaps the most potent result is that snowfall for the next 24 hours can be better predicted even by futures contracts trading up to twenty-five days in the past. Although most of our coefficients are small in magnitude, they are nevertheless significant.

The potential for information markets clearly has not been fully explored in other areas of social sciences. The ability to predict real events accurately ahead of time can present great advantages in policy decisions. It is our hope that this simple study illustrates the possibilities of utilizing information markets, so that instances of its use become more prevalent.

8. Bibliography

Roll, Richard. "Orange Juice and Weather." *The American Economic Review* 74.5 (1984): 861-80. *JSTOR*. Web. 19 Apr. 2015.

Campbell, Sean D. "Weather Forecasting for Weather Derivatives." *Journal of the American Statistical Association* 100.469 (2005): 6-16. *JSTOR*. Web. 02 Aug. 2015.

Hahn, Robert William., and Paul C. Tetlock. *Information Markets: A New Way of Making Decisions*. Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies, 2006. Print.

Cameron, Adrian Colin., and P. K. Trivedi. "20.4.1 Negative Binomial Model." *Microeconometrics: Methods and Applications*. Cambridge: Cambridge UP, 2005. 675-77. Print.

Jewson, Stephen, and Anders, Brix. "Chapter 1: Weather Derivatives and the Weather Derivatives Market." *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*. Cambridge: Cambridge UP, 2005. N. pag. Print.

Wolfers, Justin, and Eric Zitzewitz. "Prediction Markets." *American Economic Association* Spring 2004 18.2 (2004): 107-26. *JSTOR [JSTOR]*. Web. 02 June 2015. ¡http://www.jstor.org/stable/3216893¿.

Simmons, Kevin M., and Daniel Sutter. "Chapter 3: An Analysis of Tornado Casualties." *Economic and Societal Impacts of Tornadoes*. Boston: American Meteorological Society, 2011. N. pag. Print.

McKelvey, Richard D., and Peter C. Ordeshook. "Sequential Elections with Limited Information." Midwest Political Science Association 29.3 (1985): 480-512. JSTOR [JSTOR]. Web. 13 Apr. 2015.

Servan-Schreiber, Emile, Justin Wolfers, David M. Pennock, and Brian Galebach. "Prediction Markets: Does Money Matter?" Electronic Markets. Routledge, n.d. Web. 15 Apr. 2015. ¡www.electronicmarkets.org¿.Haymes (2010).

"Regression Models with Count Data." *Regression Models with Count Data*. Statistical Consulting Group, UCLA Academic Technology Services, Apr. 2007. Web. 12 July 2015. ¡http://www.ats.ucla.edu/stat/stata/seminars/count˙presentation/count.htm¿.

"Stata Data Analysis Examples: Negative Binomial Regression." *Institute for Digital Research and Education*. UCLA, n.d. Web. 10 July 2015. ¡http://www.ats.ucla.edu/stat/stata/dae/nbreg.htm¿.

"Stata Data Analysis Examples: Poisson Regression." *Institute for Digital Research and Education*. N.p., n.d. Web. 10 July 2015. ¡http://www.ats.ucla.edu/stat/stata/dae/poissonreg.htm¿.

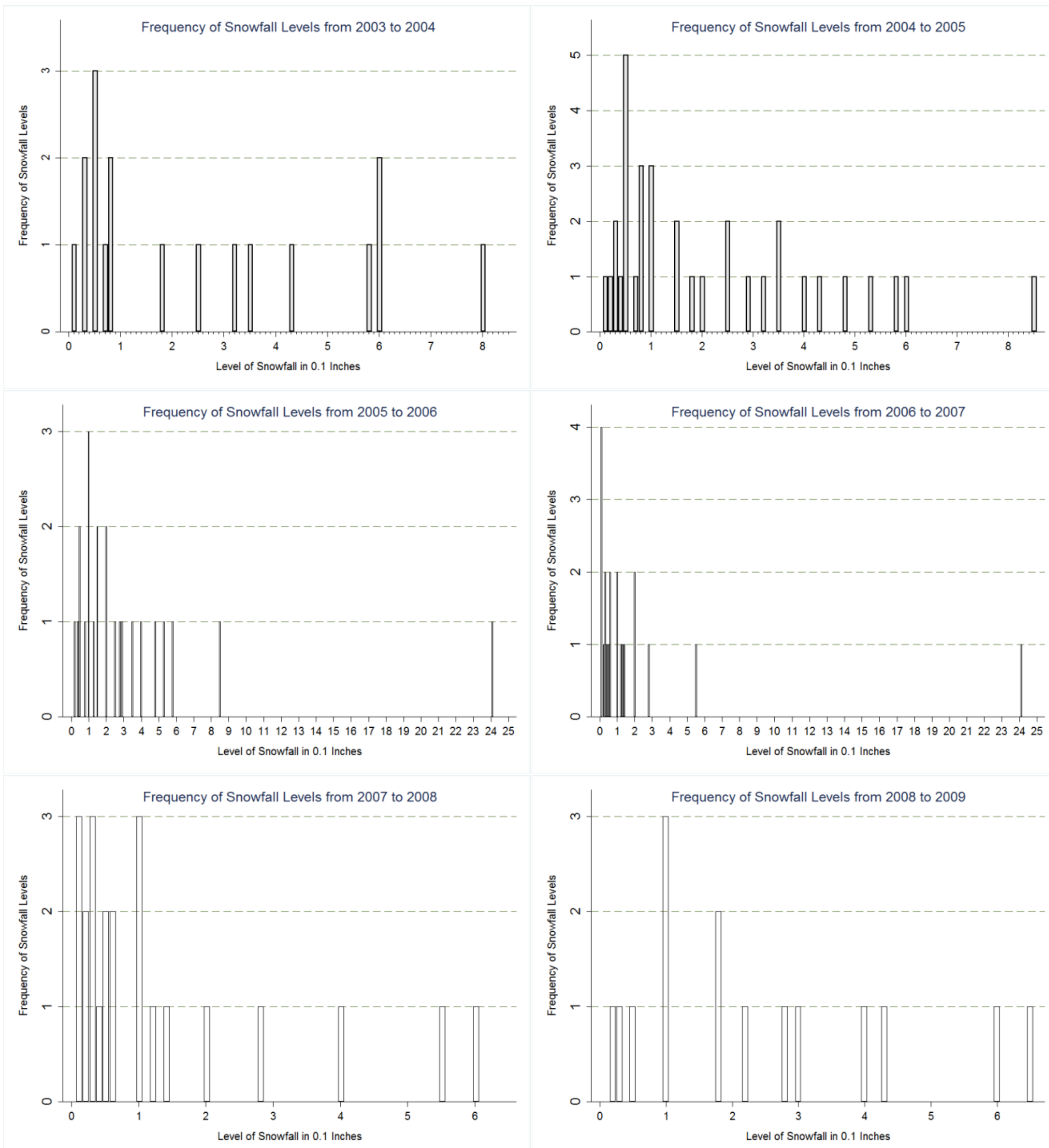
Smyth, Gordon K. "Pearson's Goodness of Fit Statistic as a Score Test Statistic." *Statistics and Science* 40 (2003): 115-26. Institute of Mathematical Statistics. Web. 15 July 2015.

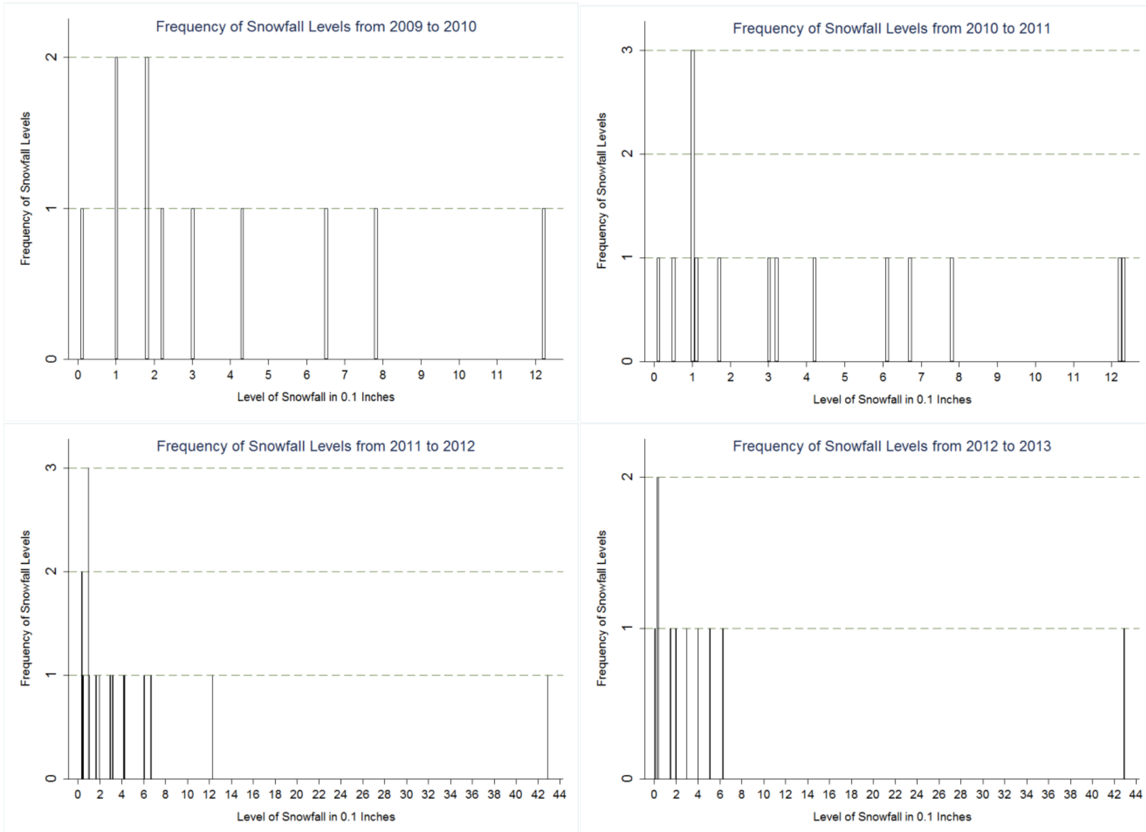
"Stata FAQ: How Can I Use Countfit in Choosing a Count Model?" *Institute for Digital Research and Education*. UCLA, n.d. Web. 03 Aug. 2015.

Nau, Robert. "How to Compare Regression Models?" *What's the Bottom Line? How to Compare Regression Models*. Fuqua School of Business, Duke University, 2015. Web. 01 Aug. 2015.

9. Figures

Figures 1a to 1j. *Annual Snowfall Conditional Distributions*

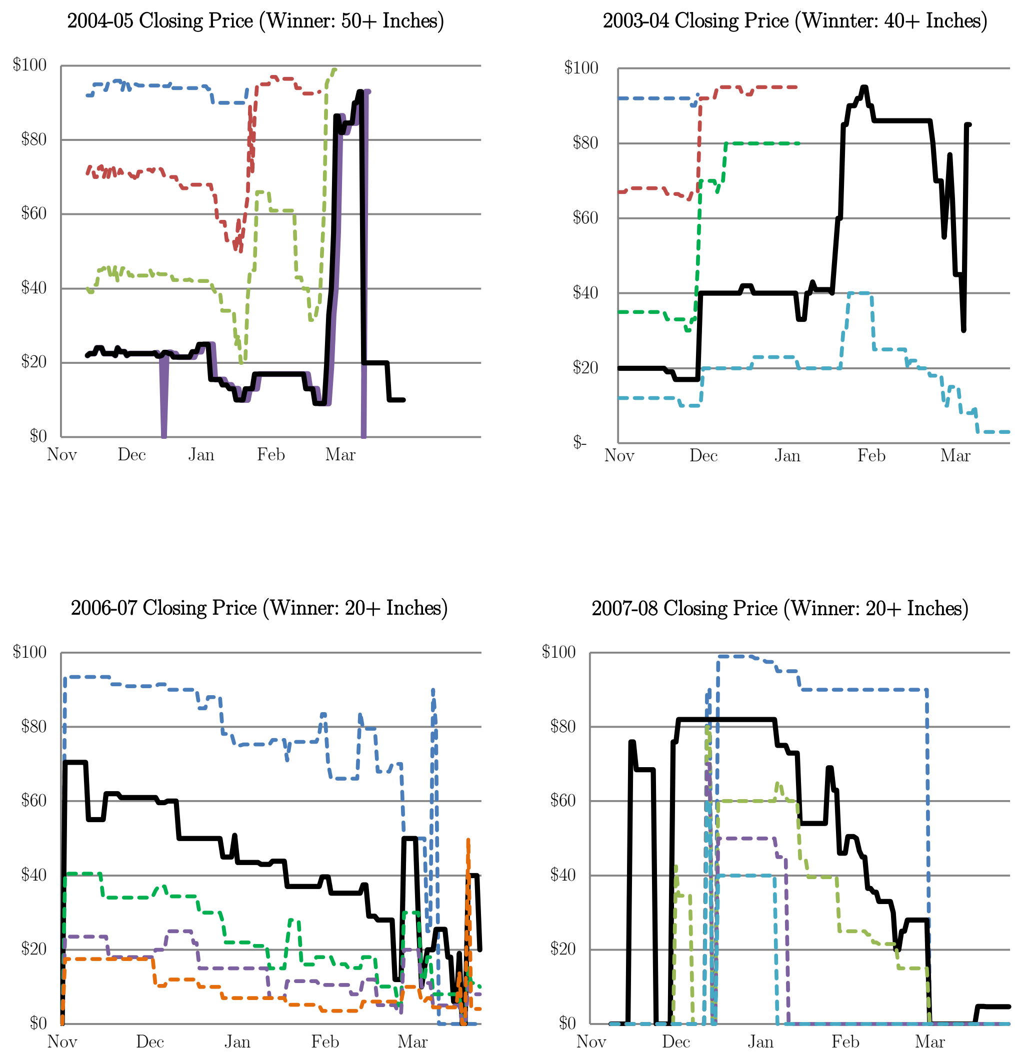



Frequency of Snowfall Levels from 2009 to 2010
Frequency of Snowfall Levels
Level of Snowfall in 0.1 Inches
Frequency of Snowfall Levels from 2010 to 2011
Frequency of Snowfall Levels
Level of Snowfall in 0.1 Inches
Frequency of Snowfall Levels from 2011 to 2012
Frequency of Snowfall Levels
Level of Snowfall in 0.1 Inches
Frequency of Snowfall Levels from 2012 to 2013
Frequency of Snowfall Levels
Level of Snowfall in 0.1 Inches

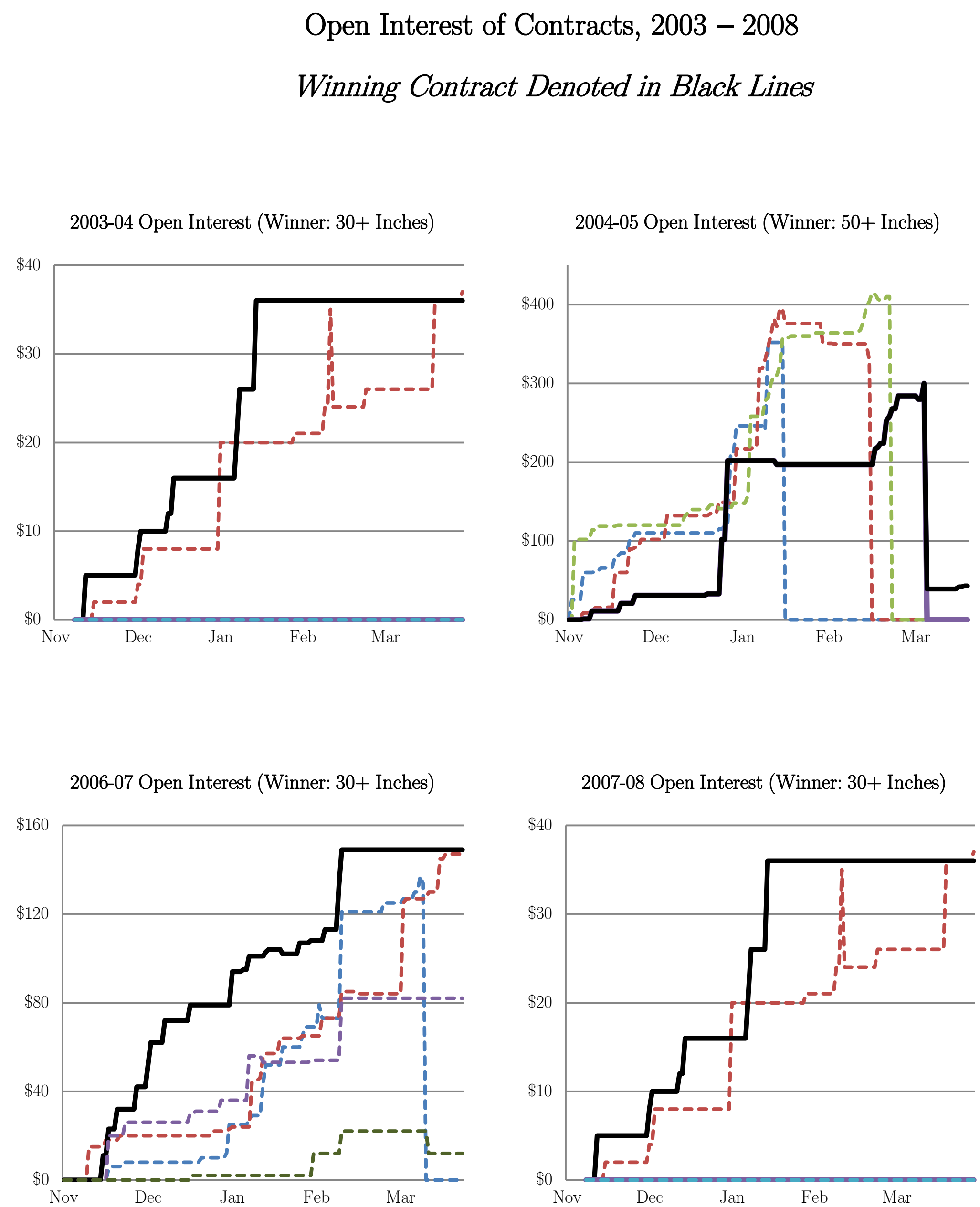
Figures 2a to 2d

Closing Price of Contracts, 2003 – 2008

Winning Contract Denoted in Black Lines



Figures 3a to 3d

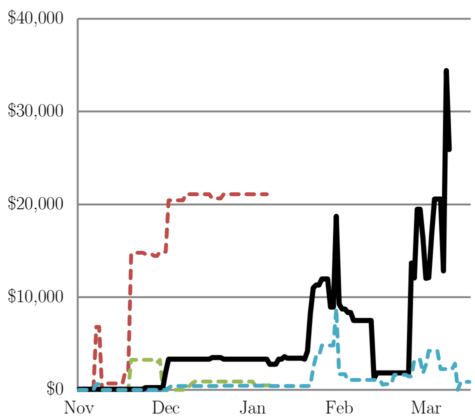


Figures 4a to 4d

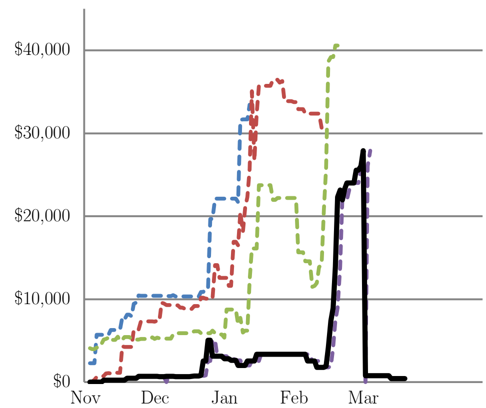
Value of Contracts, 2003 – 2008

Winning Contract Denoted in Black Lines

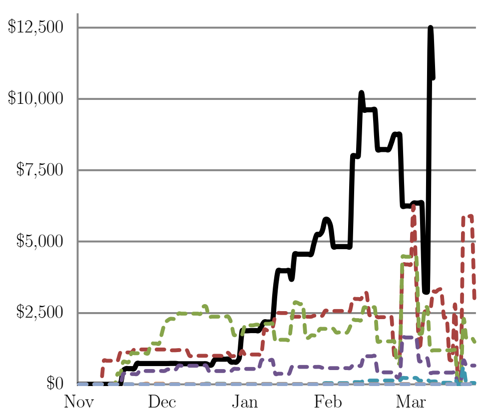
2003-04 Value (Winner: 40+ Inches)



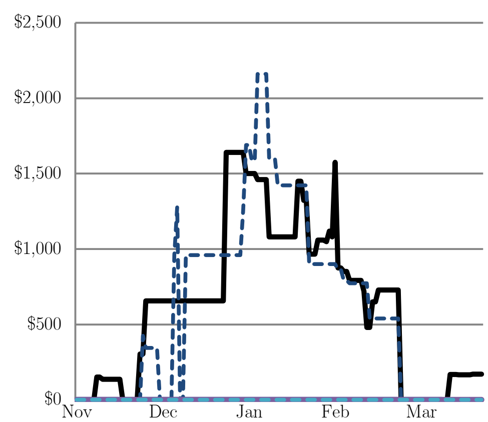
Value 2004-05 (Winner:50+ Inches)



2006-07 Value (Winner: 10+ Inches)

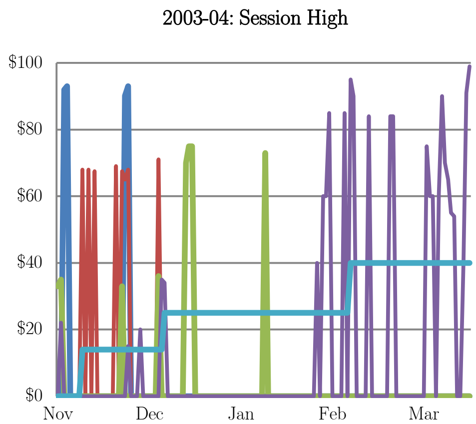


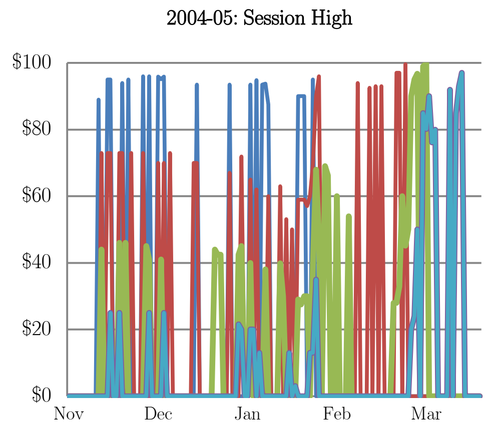
2007-08 Value (Winner: 20+ Inches)



Figures 5a to 5b

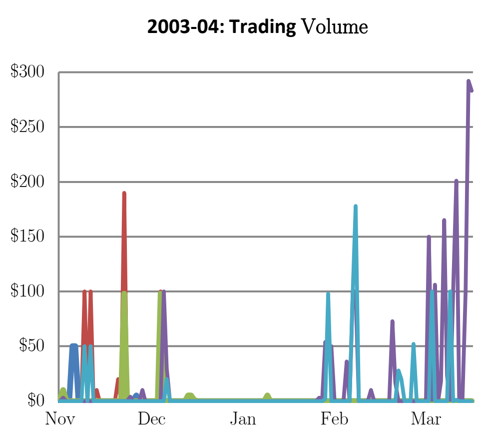
Session High of Contracts, 2003 – 2005

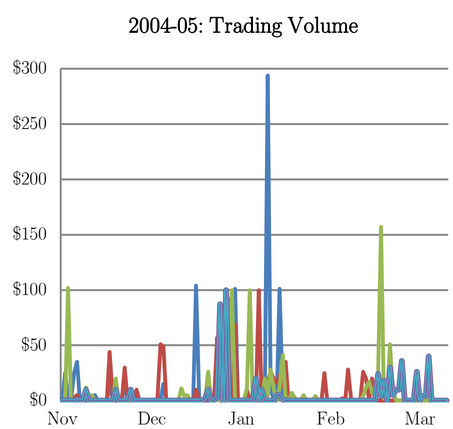




Figures 6a to 6b

Trading Volume of Contracts, 2003 – 2005

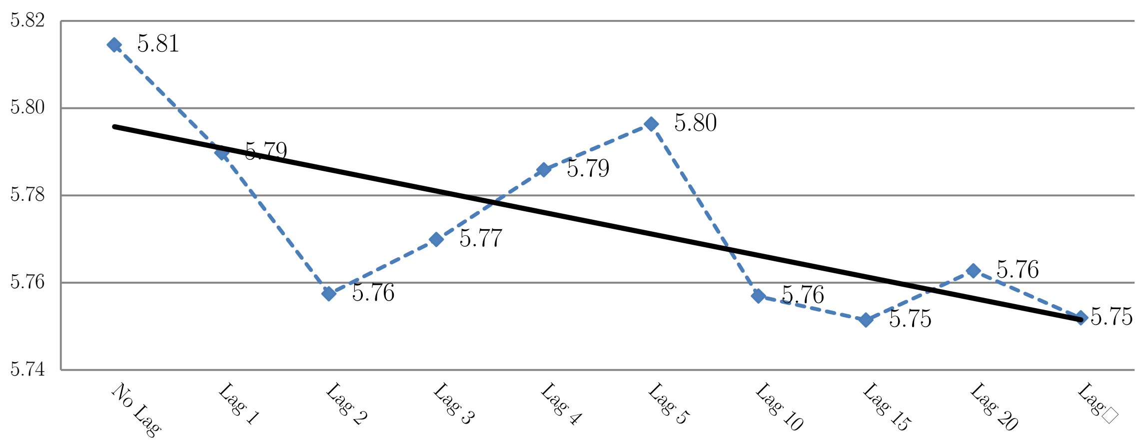




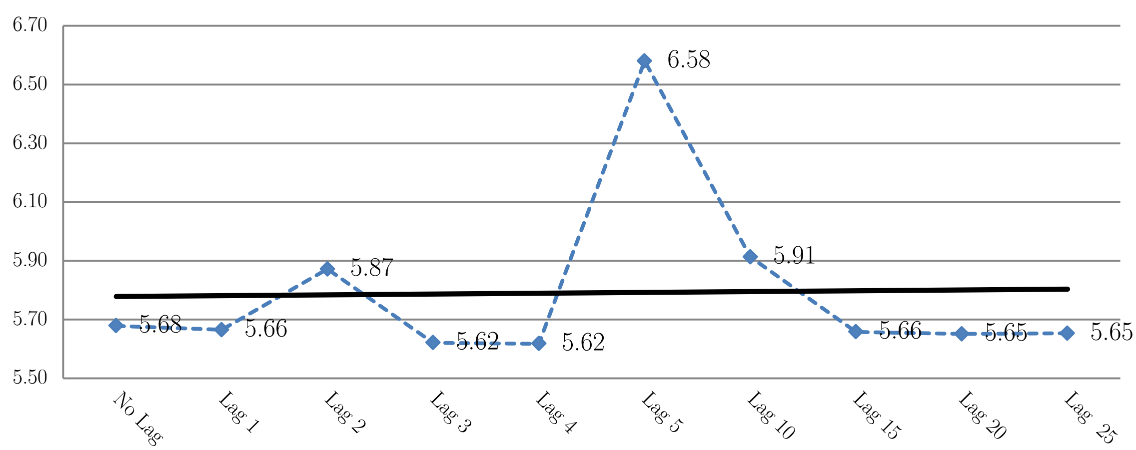
Figures 7a to 7d

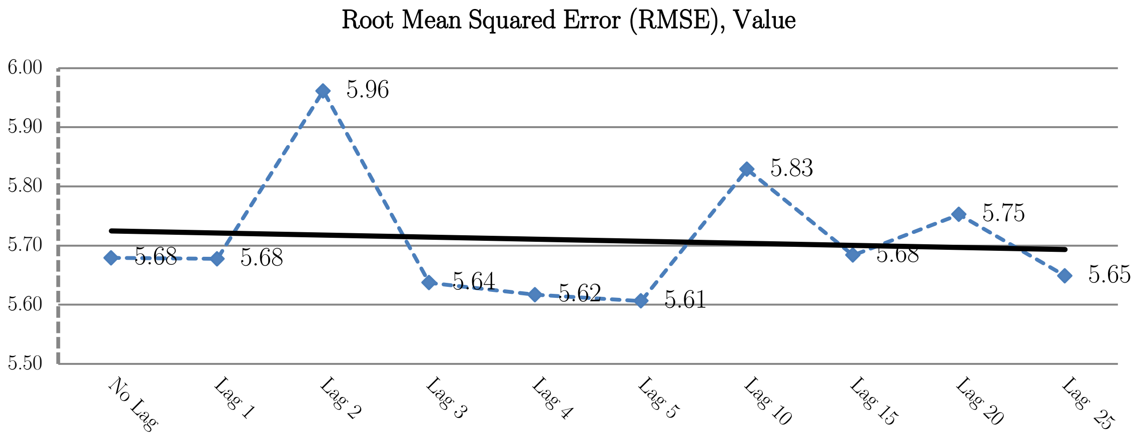
Root Mean Squared Error (RMSE) of Models

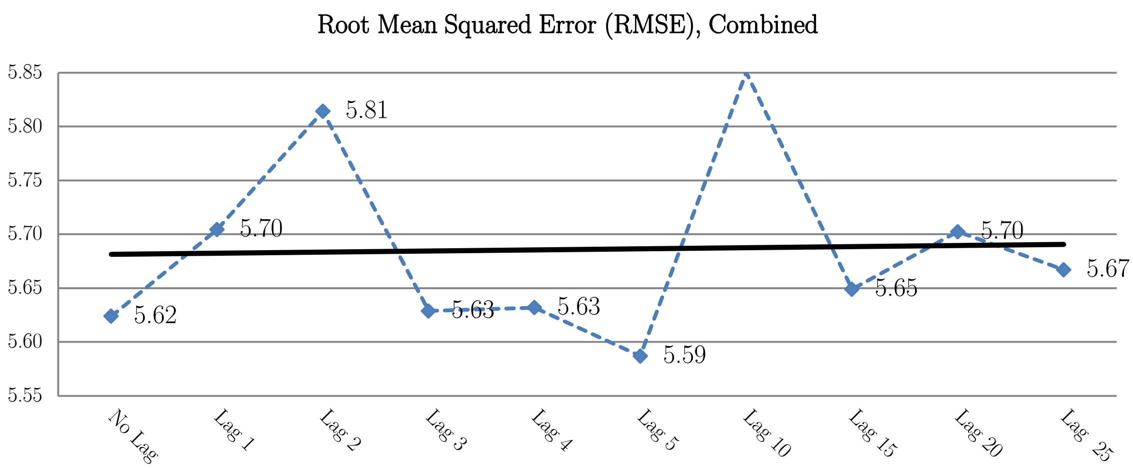
Root Mean Squared Error (RMSE), Closing Price



Root Mean Squared Error (RMSE), Open Interest




Root Mean Squared Error (RMSE), Value
6.00
5.90
5.80
5.70
5.60
5.50
5.96
5.83
5.75
5.68
5.68
5.64
5.62
5.61
5.68
5.65
No Lag
Lag 1
Lag 2
Lag 3
Lag 4
Lag 5
Lag 10
Lag 15
Lag 20
Lag 25


Root Mean Squared Error (RMSE), Combined
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.81
5.70
5.70
5.67
5.62
5.63
5.63
5.59
5.65
No Lag
Lag 1
Lag 2
Lag 3
Lag 4
Lag 5
Lag 10
Lag 15
Lag 20
Lag 25

10. Tables

Table 1 *Intrade Contract Rules for Snowfall in Central Park, NYC.*

Contract Rules: A contract will settle (expire) at 100 ($10.00) if snowfall is equal to, or greater than, the number of inches specified in the contract. A contract will settle (expire) at 0 ($0.00) if snowfall is less than the number of inches specified in the contract. Expiry will be based on snowfall data published by the National Weather Service. Any changes to the result after the contract has expired will not be taken into account - Contract Rule 1.4 Due to the nature of this prediction market contract you are obligated to read Contract Rule 1.7 (Unforeseen Circumstances) and Contract Rule 1.8 (Time Protection). Intrade may invoke these rules in its absolute discretion if deemed appropriate.

Table 2b. Frequency of Daily Snowfall Levels, December 1940-2003

	Dec 01	Dec 02	Dec 03	Dec 04	Dec 05	Dec 06	Dec 07	Dec 08	Dec 09	Dec 10	Dec 11	Dec 12	Dec 13	Dec 14	Dec 15
0.0 inches	60	59	58	58	56	56	56	59	50	57	56	56	56	56	53
0.1 inches				1	1	1				1	1		2		1
0.2 inches		1				2			2			1			
0.3 inches						1	2		1				1	1	1
0.4 inches												1			
0.5 inches			1		1				2						
0.6 inches									1	1					
0.7 inches					1						1				2
0.8 inches										1					
0.9 inches															
1.0 inches											1				
1.1 inches							2		1						
1.2 inches															
1.3 inches															
1.4 inches														1	
1.5 inches								1	1						
1.6 inches															
1.7 inches														1	
1.8 inches															
1.9 inches									1						
2.0 inches															
> 2.0 inches			1	1	1				1		1	2	1	1	3

Table 2b. Frequency of Daily Snowfall Levels, December 1940-2003

	Dec 16	Dec 17	Dec 18	Dec 19	Dec 20	Dec 21	Dec 22	Dec 23	Dec 24	Dec 25	Dec 26	Dec 27	Dec 28	Dec 29	Dec 30	Dec 31
0.0 inches	55	57	55	53	54	54	51	55	54	53	56	51	56	51	56	57
0.1 inches	1		1	1		2	1			1		1	1			
0.2 inches					1		2	2								
0.3 inches	1		1		1					1		1		3	1	
0.4 inches	1	1	1		1					2			1	1		
0.5 inches	1									1		2				
0.6 inches							1		1						2	
0.7 inches																
0.8 inches									1							
0.9 inches			1													
1.0 inches		1										1	1			
1.1 inches		1					1									
1.2 inches						1		1								
1.3 inches					1											
1.4 inches			1													
1.5 inches																
1.6 inches						1										
1.7 inches	1															
1.8 inches							1							2		
1.9 inches																
2.0 inches													1			
> 2.0 inches				6	2	2	3	2	3	2	4	3		3	1	3

Table 2c. Frequency of Daily Snowfall Levels, January 1940-2003

	Jan 01	Jan 02	Jan 03	Jan 04	Jan 05	Jan 06	Jan 07	Jan 08	Jan 09	Jan 10	Jan 11	Jan 12	Jan 13	Jan 14	Jan 15
0.0 inches	52	56	51	55	52	50	46	52	50	53	52	47	50	52	51
0.1 inches	1	1	1	1	1	2		2	1			1	1		1
0.2 inches	1					1			1	1	1	1	2	1	3
0.3 inches	1				2	2		1	1		2				1
0.4 inches		1	2			1			3			2	2		
0.5 inches	2		2		1									1	
0.6 inches		1	1	1			1				1	1			
0.7 inches			1												
0.8 inches			1						1			1			1
0.9 inches							2		1				1		1
1.0 inches							1	1	1			1			1
1.1 inches				1						1					
1.2 inches															
1.3 inches					1		2						1		
1.4 inches															
1.5 inches					1	1				1					
1.6 inches										1					
1.7 inches		1									1				
1.8 inches			1				1					2		2	
1.9 inches											1				
2.0 inches					1		3	1						1	1
> 2.0 inches	2			2		3	4	3	1	3	2	4	3	3	

Table 2d Frequency of Daily Snowfall Levels, January 1940-2003

	Jan 16	Jan 17	Jan 18	Jan 19	Jan 20	Jan 21	Jan 22	Jan 23	Jan 24	Jan 25	Jan 26	Jan 27	Jan 28	Jan 29	Jan 30	Jan 31
0.0 inches	49	52	55	52	43	53	58	55	56	51	52	51	54	53	52	49
0.1 inches	1	2			2	1				1	1	1			1	2
0.2 inches	1	1	1	2	1	1		1			1	1	1	2	1	3
0.3 inches	1			1						1		1		2		1
0.4 inches				1	1									1	2	2
0.5 inches			1		5					2						
0.6 inches	1			1						1	2		1			
0.7 inches					1		1									
0.8 inches					2					1					1	1
0.9 inches	1											1			1	1
1.0 inches			1			2		1						1		1
1.1 inches					1	1										
1.2 inches												1				
1.3 inches		3		1							1	1				
1.4 inches												1	1			
1.5 inches	1			1					1						1	
1.6 inches						2										
1.7 inches		1										1				
1.8 inches									1				1	1		
1.9 inches	1							1								
2.0 inches					2			1								
> 2.0 inches	4	1	2	1	2		1	1	2	3	3	1	2		1	

Table 2e Frequency of Daily Snowfall Levels, February 1940-2003

	Feb 01	Feb 02	Feb 03	Feb 04	Feb 05	Feb 06	Feb 07	Feb 08	Feb 09	Feb 10	Feb 11	Feb 12	Feb 13	Feb 14	Feb 15
0.0 inches	53	48	53	50	54	47	50	48	53	57	52	50	48	47	49
0.1 inches	1	2		3		1					2		2		3
0.2 inches		2		1	1		1	2					2		1
0.3 inches			1	1		2			1			1		1	
0.4 inches		1				2	2		2				1		2
0.5 inches		1	2	1				3			1	1			
0.6 inches	1		1										1		
0.7 inches															
0.8 inches			1											2	
0.9 inches	1					1						1			
1.0 inches		2	1					1					1	1	1
1.1 inches												1	1	2	
1.2 inches									1					1	
1.3 inches										1			2		1
1.4 inches							1								
1.5 inches		1													
1.6 inches		1													
1.7 inches											1	2			
1.8 inches				1					1						
1.9 inches						1									
2.0 inches	2					1				1			1	1	1
> 2.0 inches	2	2	2	3	5	5	6	5	2	1	4	4	1	5	2

Table 2f Frequency of Daily Snowfall Levels, February 1940-2003

	Feb 16	Feb 17	Feb 18	Feb 19	Feb 20	Feb 21	Feb 22	Feb 23	Feb 24	Feb 25	Feb 26	Feb 27	Feb 28
0.0 inches	50	55	54	53	55	56	52	52	53	51	55	57	53
0.1 inches									1		2		
0.2 inches		1				1				1			
0.3 inches	1						2		2	2			1
0.4 inches	1		1		1	1	1	2	1	1	1		1
0.5 inches			1				1				1		1
0.6 inches	1			1								2	
0.7 inches	2		1						1				
0.8 inches	1	1											1
0.9 inches													1
1.0 inches	1							1	1				
1.1 inches							1						
1.2 inches													
1.3 inches													
1.4 inches		1								1	1		
1.5 inches		1			1					1			
1.6 inches	1						2	1					
1.7 inches		1	1										
1.8 inches									1	2			
1.9 inches					1								
2.0 inches					1							1	1
> 2.0 inches	2		2	6	1	2	1	4		1			1

Table 2g Frequency of Daily Snowfall Levels, March 1940-2003

	Mar 01	Mar 02	Mar 03	Mar 04	Mar 05	Mar 06	Mar 07	Mar 08	Mar 09	Mar 10	Mar 11	Mar 12	Mar 13	Mar 14	Mar 15
0.0 inches	51	53	51	55	54	55	54	52	58	56	59	53	53	50	58
0.1 inches	3	1		1	2			4							
0.2 inches				1	1	1	2			1		1		1	
0.3 inches	1				1		1			1		2	1	1	
0.4 inches													1	1	
0.5 inches										1				1	
0.6 inches								1		1					
0.7 inches				1									1		
0.8 inches	1														
0.9 inches												1			
1.0 inches			3								1			1	
1.1 inches												1		1	
1.2 inches					1		1								
1.3 inches														1	
1.4 inches		1	1										1	1	
1.5 inches															
1.6 inches												1			
1.7 inches			1												
1.8 inches		1						1							1
1.9 inches															
2.0 inches	1		1	2	1										
> 2.0 inches	3	4	3			4	2	2	2			1	3	1	1

Table 2h. Frequency of Daily Snowfall Levels, March 1940-2003

	Mar 16	Mar 17	Mar 18	Mar 19	Mar 20	Mar 21	Mar 22	Mar 23	Mar 24	Mar 25	Mar 26	Mar 27	Mar 28	Mar 29	Mar 30	Mar 31
0.0 inches	55	52	55	56	56	56	53	60	59	60	57	59	57	54	59	58
0.1 inches			1													2
0.2 inches	1		1			1					2		1	1		
0.3 inches		1											1			
0.4 inches		1										1				
0.5 inches		1													1	
0.6 inches			1		1						1					
0.7 inches																
0.8 inches	1	1				1										
0.9 inches																
1.0 inches																
1.1 inches		1														
1.2 inches									1							
1.3 inches						1										
1.4 inches																
1.5 inches		2											1			
1.6 inches																
1.7 inches					1											
1.8 inches	1													1		
1.9 inches							2									
2.0 inches	1			1												
> 2.0 inches	1	1	2	3	2	1	5							4		