PRICING AND PROBABILITY DISTRIBUTIONS OF ATMOSPHERIC VARIABLES

TECHNICAL WHITE PAPER

WILLIAM M. BRIGGS

Abstract. Current methods of assessing the probability distributions of atmospheric variables are inadequate. These distributions are the key input to pricing models of weather derivatives and they are used for assessing the cost of insurance policies written against the atmospheric variables. Inaccurate inputs to these models lead to incorrectly priced derivatives and risky policies, which of course lead companies to lose money.

It is possible to provide better and more accurate probability distributions for atmospheric variables based on advances in a wide-range of meteorological, climatological, and statistical research. This paper contrasts older technologies with what can be offered and it gives several examples of how dramatic a difference an improvement in estimation techniques can make to the bottom line. These improvements can, if used wisely, lead directly to increased profits for your company.

1. Basic Methodology

The atmospheric variable most frequently used for pricing weather derivatives and insurance policies is temperature. Heating and Cooling Degree Days (HDD, CDD) are used in derivatives, and daily maximums, minimums, and averages (to list only a few) are used in insurance. The following methods can be used with any variable, but since temperature is ubiquitous it will be used as our primary example.

When companies fix the price of a temperature derivative or set the policy price on a temperature they frequently consider only temperature in isolation, ignoring the effects of other atmospheric variables. A simple example will indicate how only considering temperature leads to the setting of (possibly wildly) incorrect prices.

Figure 1 shows (in blue) the density for average daily temperature for New York City in January[1]. As expected, the density is normal and has a mean of 36°F. But look at what happens to the density on those days on which measurable precipitation falls (red line). The density markedly shifts to warmer temperatures (mostly because of increased nighttime cloud cover and increased daily lows). Similarly, (shown in green) days on which no precipitation falls show a shift toward cooler temperatures.

These shifts in temperature effect only (or mostly) the tails of these densities, as a glance at Table 1 will confirm. The average temperature for both the Unconditional and Precipitation days is 36°F, falling only one degree to 35°F on Non Precipitation days. But the presence of precipitation dramatically influences probability estimates for extreme temperatures. A similar effect holds for the densities

[1]Data from National Climatic Data Center, 1997-2001.

of monthly and seasonal HDDs and CDDs under the influence of, among other things, the El Niño Southern Oscillation (ENSO).

> **"(It is) important to estimate the changes (in probability distributions of atmospheric variables) accurately for each ENSO event, because even small changes of means and variances can imply large changes of the likelihood of extreme values."** —Saradeshmuch, Compo, and Penland. Journal of Climate.

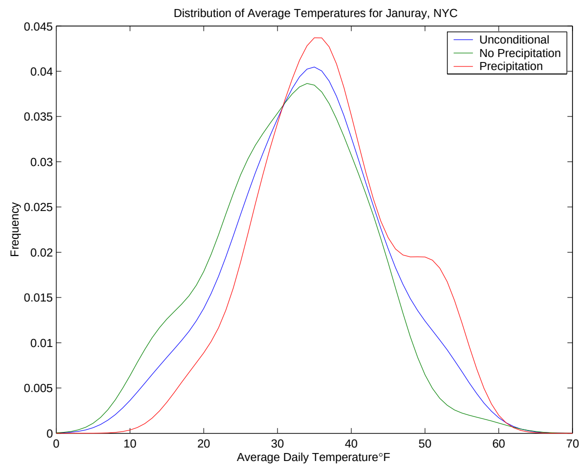

FIGURE 1. Average daily temperature distribution for New York City in January. The Unconditional (blue) line indicates the distribution most firms would use. The No Precipitation (green) and Precipitation (red) lines indicate how this distribution shifts in the presence or absence of precipitation.

SCENARIO ONE: INSURANCE

Suppose an insurance company has been asked to write a policy which will pay $10,000.00 if the average temperature on January 15th falls below 20°F. Unconditionally, the probability for this is about 10% (see Table 1). This estimate allows the underwriter to calculate the expected payout the insurance company would face, which in this case is about $1000.00. This number will be the basis in setting the price of the policy (reflecting the base price before the addition of the risk premium).

But what if the underwriter knew for certain that precipitation would not fall on January 15? The estimate for below 20°F now jumps to 15%, which boosts the expected payout to $1500.00, a major difference.

Similarly, if it was known that precipitation *would* fall the estimate becomes 4% for an expected payout of only $400.00, also a large difference from the unconditional payout.

If the underwriter does not take precipitation into account he could set the policy price too high and lose business to competition, or set it too low and face

TABLE 1. The average temperature and probability of the daily average falling below 20°F for New York City in January for the average across all days (Unconditional), days on which no precipitation fell (Dry) and days on which measurable precipitation fell (Wet).

	Average Temperature	Prob. $T < 20$°F
Unconditional	36	0.10
Dry	36	0.15
Wet	35	0.04

unnecessary risk. It is crucial, therefore, to understand how temperature interacts with other atmospheric variables.

NEW METHOD

One important, and often overlooked, factor is the weather or climate forecast for the relevant atmospheric variables. These forecasts are provided by, among others, the National Weather Service (NWS) and the Climate Prediction Center (CPC). In actual situations the contract writer would not know for certain that precipitation would or would not fall and would therefore not know how to adjust his base probability estimate. He may be aware of the forecast for temperature and for precipitation (but unaware of the forecast for other related variables). But he won't know how to use these forecasts.

There are proven peer-reviewed methods that incorporate forecast information and that allow the adjustment of base probability estimates. It is important to also know when to look at a variable in isolation or in conjunction with other variables. These methods work on any time scale (daily, weekly, monthly, yearly, or odd combinations) and with any variable (temperature, heating and cooling degree days, rain or snow, sunshine or cloud cover, and so on). The above example will be continued to demonstrate the methods.

SCENARIO TWO: INSURANCE

Suppose, for the sake of simplicity, that the forecast for the average temperature (which is derived from the forecast for the high and low temperature forecasts) for January 15 is for 36°F, i.e. the climatological average. We know from the data that the climatological probability of precipitation is about 46%. If the forecast for precipitation is 46% we would use (for average temperature below 20°F) the unconditional estimate of 10%. If the forecast for precipitation is 0% or 100% we would use either the estimates of 15% or 4% as before. But suppose the forecast is for a 10% chance or precipitation. New methods show that this modifies the base probability to 14%, slightly lower than the precipitation free estimate as expected, and still an important departure from the unconditional estimate.

SCENARIO ONE: DERIVATIVE

The same analysis can be performed for a monthly HDD weather derivative. Here, direct probabilities are issued which forecast the possibility that the monthly or seasonal average temperatures and precipitation totals will be at, below, or above normal. These probabilities can be used directly to modify estimates for distributions of any atmospheric variable. We show here that the forecast for monthly total precipitation also affects the distribution of HDDs.

Suppose an energy company would like to hedge against warmer than average temperatures for January in New York City. The mean[2] number of HDDs is 1002. The energy company wants to be paid \$5000.00 for every HDD under 800. So, for example, if the actual number of HDDs turned out to be 750, the energy company would be paid $(800-750)\times\$5000.00 = \$25,000.00$. Using a nonparametric estimate of the unconditional distribution for HDD we find that the expected value of the payout is about \$64,900.00.

Now suppose that the CPC issues a maximum forecast for above average precipitation. The effect this has on the distribution of HDDs is shown in Fig. 2. The solid blue line is the unconditional density for HDDs. The broken red line is the density conditional on the precipitation forecast. As expected, because of the increased chance of precipitation the number of HDDs has decreased reflecting the possible warmer temperatures. This change results in the expected payout to increase to \$69,600.00, a difference of 7%. Clearly, someone can benefit from the incorrect pricing of the original derivative if they have information about the precipitation forecast.

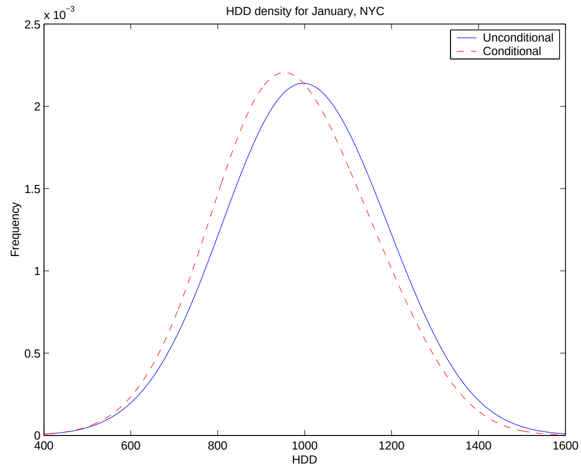

Figure 2. HDD densities under normal and wet conditions.

SCENARIO THREE: INSURANCE

We can now complicate the situation for the insurance underwriter by adding the forecast for temperature. Suppose the forecast for the high temperature is 38°F and the forecast for the low temperature is 26°F. The climatological averages are 40°F and 28°F. Keep the forecast for precipitation at 10%. New methods show that these forecasts modify the base probability to anywhere from 0% to as much as 14%. Why? Because of the range of and the uncertainty of the temperature forecasts. If the forecasts for high and low temperatures were 100% certain then the daily average would be exactly $(38+26)/2 = 32$°F, which implies the probability of falling below 20°F is zero (making the impact of the precipitation forecast unimportant). As the forecast becomes more uncertain, that is, that the high and low won't exactly

[2]We arbitrarily used the period 1950-2001 to compute the mean, although the period of reference for real contracts is chosen by more careful methods.

be 38°F or 26°F, it becomes more probable that the average will fall below 20°F. The next section explains how we find and use this uncertainty.

Emphasis must be given, in this example, that even though the temperature forecast is for lower than the climatological average the probability of the daily average being below 20°F does not necessarily rise: it can fall quite dramatically.

SCENARIO TWO: DERIVATIVE

We can incorporate temperature forecasts for derivative pricing too. Suppose the CPC issued a joint forecast for maximum above average monthly temperature and monthly precipitation. The payout now rises to a dramatic $93,700.00. A joint minimum forecast for minimum average monthly temperature and monthly precipitation gives an expected payout of only $23,500.00. The real payout will lie somewhere in the range of [$23,500.00, $93,700]. New methods can show you where the best bet is.

These examples have ignored many important factors, such as trends in historical data, prediction of values at locations other than weather stations, corrupted data, and many others. All of these factors can be accounted for in the methodology we have developed. Questions about a specific factor are welcomed.

2. Uncertainty and Past Forecasts

Uncertainty for a precipitation forecast exists when the forecast is given in the form of a probability, which is always done by the NWS. But this isn't true for temperature forecasts as normally seen by the public who only get a single number (which implies a certainty that certainly isn't there). There are two different notions of confidence given in probabilistic daily temperature forecasts and in probabilistic climatic temperature forecasts. These may be quantified, and are, by using both the NWS and CPC publically available data.

For daily temperature forecasts the NWS issues point (single number) forecasts that can be, using new methods, turned into central credible interval (CCI) forecasts. A CCI is a forecast of a range of values and a specific probability that the actual value will fall within this range. For example, a daily high temperature 50% CCI forecast of [40°45°] says that there is a 50% chance that the high temperature will fall between (and including) 40° and 45°, and a 50% chance that the actual high will be less than 40° or more than 45°. Both the probability (here it was 50%) and the size of the interval can be varied to fully specify the uncertainty[3].

Monthly and seasonal average temperatures are handled differently. Here, direct probabilities are issued by the CPC that forecast the possibility that the monthly or seasonal average temperatures will be at, below, or above normal. These probabilities can be used directly to modify estimates for distributions of any atmospheric variable. Examples will be given on request, but intuitively , as the probability that the seasonal average temperature will be above normal rises this changes the distributions of seasonal temperature, shifting it towards warmer values (*important:* this forecast also modifies the distributions for other atmospheric variables too). The techniques developed to use these forecasts are especially useful for monthly and seasonal CDD and HDD weather derivative contracts.

[3]Different forms of uncertainty are possible: for example, the entire distribution of possible temperatures may be given. These distributions provide more information than CCI forecasts but are harder to construct.

There is another notion of confidence that the actual forecasts do not but should include, *and it is one that you can use in forecasts of any type, not just atmospheric.* This notion is that past forecast accuracy is predictive of future forecast accuracy. This field is vast, and completely untapped by the business world. You should be aware that by not taking advantage of the tools of forecast evaluation you are either losing money or not making as much as you should.

CALIBRATION

We first examine calibration using the example of precipitation forecasts.

Suppose we have a collection of precipitation forecasts[4] and the accompanying observations. For the set of times the forecaster gave a probability of 0.10 a calibrated set of forecasts would have 10% of those occasions resulting in precipitation. Likewise, each time the forecaster said 0.70 we would expect that 70% of those occasions that precipitation fell. Full calibration demands that for each unique forecast number issued by the forecaster that same percentage of times precipitation should have occurred. A graph will help simplify the meaning.

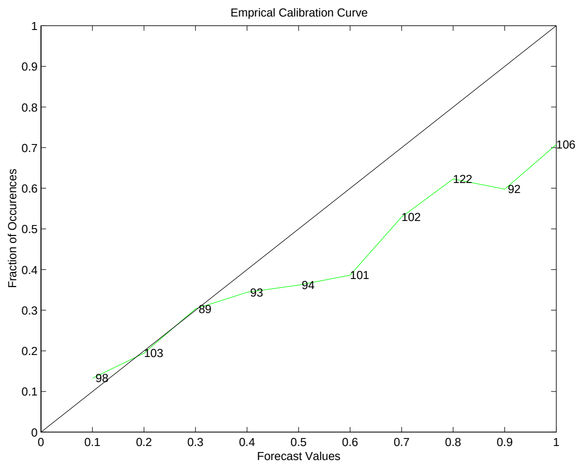

FIGURE 3. Calibration Curve. See the text for its interpretation.

A curve of this shape (Fig. 3) is very common for forecasts generated by a person (and not a computer algorithm): it expresses overconfidence in the event. The horizontal-axis lists the forecast values, here 0.1, 0.2, and so on. The numbers at each forecast value show the number of times that forecast value was used. For example, the forecast value of 0.5 was used 94 times. In this example all forecast values are used roughly equally (with the exception of 0). The vertical-axis shows the frequency of occurrences of the event given the forecast value. Perfectly calibrated forecasts would line up along the one-to-one line in the center of the graph. The green line shows the actual state of affairs for a certain precipitation forecaster.

The overconfidence is apparent. Take the forecast value of 0.9. Only 60% of these times did precipitation occur, meaning that the forecaster was more certain

[4]They could be forecasts for any dichotomous event: price rise or fall, a type of sale, and so on.

than he should have been. The forecaster does do a good job with forecast values of 0.3 or less; he appears to be calibrated at these values.

How does calibration fit in with the methods given in the first section? Obviously, uncalibrated forecasts aren't as good as calibrated ones[5], so that if we blindly used the forecast without assessing its calibration we could be missestimating the probability distributions we are interested in.

SCENARIO FOUR: INSURANCE

Suppose, as before, the forecast for the daily temperature reflects climatology and that the forecast for precipitation given by our forecaster is 0.90 (the same forecaster responsible for Fig. 2). As before we're interested in the probability that the average temperature on January 15 falls below 20°F. Using our method without regard to calibration gives a probability of 5%.

But if we use a *recalibrated* forecast the estimate increases to 9%, a large difference. Clearly, calibration is extremely important.

Using the recalibrated forecast carries the implication that the forecaster will be uncalibrated to the same (statistical) extent that he has been in the past. This need not be the case, of course, because tools such as Fig. 3 can be used to improve future performance.

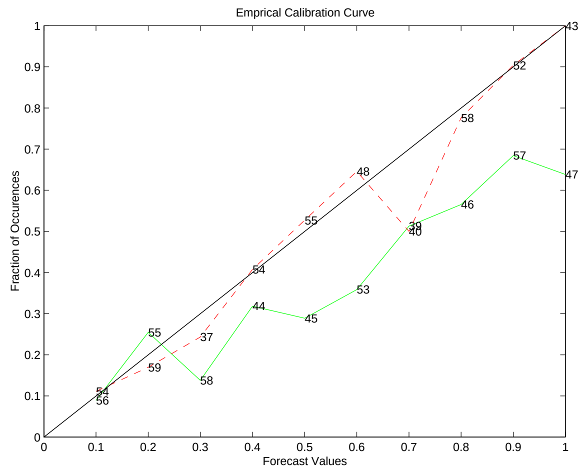

FIGURE 4. Calibration Curve broken into two regimes: one in which the high temperature was greater than 50°F (solid green line), and one in which the high temperature was less than 50°F (broken red line).

For example, Fig. 4 shows the calibration curve for the same forecaster, only this time the forecasts have been split into two regimes, one in which the high temperature was greater than 50°F (solid green line), and one in which the high temperature was less than 50°F (broken red line). In practice, any external variable might be influential—*new methods allow us to determine which ones and by what extent*—temperature is used only as an example.

---

[5] It can be *proved* that calibrated forecasts are *always* better than uncalibrated forecasts no matter what the loss function is.

The forecaster appeared uncalibrated before, but now it shows that when the high temperature is less than 50°F he is very nearly calibrated, and that his overconfidence is associated with warm temperatures. If the forecaster is presented with this information he should be able to correct his biases.

There are many aspects of verification not touched upon here: reliability or sharpness, continuous valued variables, skill, and more. Other white papers will address these.

300 E. 71st, Apt. 3R, New York, NY, 10021
E-mail address: william@wmbriggs.com