View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

NCER National Centre for Econometric Research

NCER Working Paper Series

Developing analytical distributions for temperature indices for the purposes of pricing temperature-based weather derivatives

A E Clements, A S Hurn, K A Liindsay

Working Paper #34

September 2008

Developing analytical distributions for temperature indices for the purposes of pricing temperature-based weather derivatives

A E Clements, A S Hurn
School of Economics and Finance, Queensland University of Technology

K A Lindsay
Department of Mathematics, University of Glasgow.

Abstract

Temperature-based weather derivatives are written on an index which is normally defined to be a nonlinear function of average daily temperatures. Recent empirical work has demonstrated the usefulness of simple time-series models of temperature for estimating the payoffs to these instruments. This paper develops analytical distributions of temperature indices on which temperature derivatives are written. If deviations of daily temperature from its expected value is modelled as an Ornstein-Uhlenbeck process with time-varying variance, then the distributions of the temperature index on which the derivative is written is the sum of truncated, correlated Gaussian deviates. The key result of this paper is to provide an analytical approximation to the distribution of this sum, thus allowing the accurate computation of payoffs without the need for any simulation. A data set comprising average daily temperature spanning over a hundred years for four Australian cities is used to demonstrate the efficacy of this approach for estimating the payoffs to temperature derivatives. It is demonstrated that expected payoffs computed directly from historical records is a particulary poor approach to the problem when there are trends in underlying average daily temperature. It is shown that the proposed analytical approach is superior to historical pricing.

Keywords
Weather Derivatives, Temperature Models, Cooling Degree Days, Maximum Likelihood Estimation, Distributions for Correlated Variables.

JEL Classification Numbers
C14, C52.

Corresponding author
Kenneth Lindsay
Department of Mathematics
University of Glasgow
Glasgow, G12 8QW
United Kingdom

email `kal@maths.gla.ac.uk`

1 Introduction

There has been growing interest in weather derivatives that permit the financial risk associated with climatic conditions such as temperature or rainfall to be managed. Similar to the situation in financial markets where a derivative security takes its value from an underlying financial asset or index, a weather derivative takes its value from an underlying measure of weather, such as temperature, rainfall or snowfall over a particular period of time. The first weather derivative was transacted in the US in 1996 and the size of the market is now in excess of US$ 8 billion.[1] Because temperature and precipitation intrinsically cannot be traded, there is no arbitrage-free pricing framework available to price these weather derivatives. Consequently this paper is primarily concerned with the development of accurate estimates of the expected payoffs of weather derivatives which is the crux of any pricing strategy.

Despite the existence of precipitation-based derivatives, the vast majority of all weather derivatives are based on temperature indices, such as heating degree days and cooling degree days.[2] Temperature derivatives are currently written on temperature indices collected from several US and European cities as well as two Japanese cities. Major participants in this market include utilities and insurance companies along with other firms with costs or revenues that are dependent upon temperature. For example, an electricity supplier normally provides its customers with electricity at a fixed price irrespective of the wholesale price. On the other hand the wholesale price of electricity can fluctuate wildly with extreme temperatures, and so temperature-based derivatives can provide a hedging tool for fluctuations in wholesale electricity prices. Consequently the focus of this paper will be exclusively on temperature-based derivatives.

The most straightforward of estimating expected payoffs is from historical records (Zeng, 2000; Platen and West, 2003). A more elaborate method is to fit a model to the time-series of average temperature so as to capture seasonal variations in both temperature and its volatility (Platen and West, 2003; Campbell and Diebold, 2004). The model is then used to simulate temperature outcomes over the period of the contract in order to construct the distribution of the temperature-based index on which the derivative is written. Note that widely-available meteorological forecasts are not suitable for this purpose because these forecasts are made over relatively short horizons, such as 7 days, whereas temperature derivatives are often traded well before[3] contracts generate any payoffs (Wilks, 1995; Jewson and Caballero, 2003; Campbell and

[1] The first recorded activity was an over-the-counter heating degree day swap option between Entergy-Koch and Enron for the winter of 1997 in Milwaukee, Wisconsin (Tindall 2006).

[2] Garmen *et al.*, 2000 posit that 98-99% of all weather derivatives currently traded are based on temperature.

[3] For example, participants in temperature-based weather derivative may enter into a contract many months before the arrival of the summer on which the payoff of the contract is to be determined.

Diebold, 2004).

Relatively few attempts have been made at generating closed-form expressions for the expected payoffs to temperature-based derivatives (Benth and Šaltynė-Benth, 2005). The fundamental contribution of this paper is to develop closed-form approximations to the distributions of the indices on which temperature-based derivatives are written. This is necessarily a complex task given that the relevant indices are nonlinear functions of average daily temperature in that they form a sequence of correlated, truncated random variables in which the level and frequency of truncation is not negligible. The basis of the analysis is the assumption that the deviations of average daily temperature from its expected value behaves as an Ornstein-Uhlenbeck process with time-varying variance. One of the primary tools used in establishing the results presented in the paper is that Riemann-Stieltjes integrals of Gaussian processes are themselves Gaussian processes and consequently the distributions of the indices on which the temperature derivatives are written are essentially the sum of correlated, truncated Gaussian distributions.

For the empirical work in this paper a data set comprising average daily temperatures for over a century in four Australian cities, namely, Brisbane, Melbourne, Perth and Sydney was collected. These locations were chosen primarily because they are the four major cities of Australia, and also because accurate temperature records of long-duration are available at single weather stations, an important institutional requirement for writing temperature-based derivatives. This is a quality data set which represents a substantial improvement on what appears to be the current standard used in the literature. The potential downside of using Australian temperature data is that Australia currently has no organized market for temperature derivatives such as that organized by the Chicago Mercantile Exchange (CME) or the London International Financial Futures and Options Exchange (Liffe).[4] Consequently, no actually observed derivative prices can be used in this analysis. Nevertheless, the methodology developed here is generally applicable and could be used to estimate the payoffs to temperature derivatives in any market.

The rest of the paper is structured as follows. Section 2 describes the data used in this investigation. Section 3 outlines the concept of the 'tick value' of a temperature-based derivatives and the importance of expected payoff in its pricing. Section 5 presents a simple continuous-time autoregressive model average daily temperature and describes how the parameters of the model may be estimated. Analytical distributions for the relevant temperature index on which derivatives are written developed in Section 5 and the use of these distributional results are demonstrated in practice in Section 6. Section 7 is a brief conclusion.

[4]Trading of weather derivatives on the CME began in September 1999 and by 2006 approximately 55% of all weather derivative trading was transacted on the CME. By contrast, in 2004 Liffe started trading weather derivatives in July 2001 but suspended trading in these instruments in 2004 due to a lack of turnover (Tindall 2006).

2 Data

The data set comprises daily maximum and minimum temperature records in degrees Celsius for Brisbane, Melbourne, Perth and Sydney.[5] Following standard practice in pricing weather derivatives (Zeng, 2000; Platen and West, 2003; and Campbell and Diebold, 2004), the analysis is conducted on the time series of average daily temperatures computed as the arithmetic mean of the daily maximum and minimum values. For all the data sets, instances of single missing values were treated by averaging adjacent records. In a few rare cases where several days were missing, the long term average for those days was inserted. Finally, following Campbell and Diebold (2004), all occurrences of the 29 February were removed.

Brisbane, Melbourne, Perth and Sydney were chosen primarily because they are the four major cities of Australia, and also because accurate temperature records of over 100 years are available for these cities at comparable weather stations. The construction of the temperature record for each city is now discussed in more detail.

Brisbane The temperature record contains 44043 observations starting on the 1/1/1887 and ending on 31/8/2007. The time series is constructed from data collected from three weather stations: Brisbane Regional Office (Station Number 40214) 1/1/1887 - 31/3/1986; Brisbane Airport (Station Number 40223) 1/4/1986 - 14/2/2000); and again from Brisbane Airport (Station Number 40842) 15/2/2000 - 31/8/2007.

Melbourne The temperature record contains 55358 observations starting on 1/1/1856 and ending on 31/8/2007. The time series is a continuous set of observations made at the Melbourne Regional Office (Station Number 86071) weather station. The location of the office changed in the early 1980s although the name of station did not.

Perth The temperature record contains 40393 observations starting on 1/1/1897 and ending on 31/8/2007. The time series is constructed from data collected at two weather stations: Perth Regional Office (Station Number 9034) 1/1/1897 - 2/6/1944; and Perth Airport (Station Number 9021) 3/6/1944 - 31/8/2007.

Sydney The temperature record contains 54263 observations starting on 1/1/1859 and ending on 31/8/2007. The time series is a continuous set of observations made at the Sydney Observatory Hill (Station Number 66062) weather station.

[5] All the raw data were supplied by Climate Information Services, National Climate Centre, Australian Bureau of Meteorology.

Summary statistics for the average daily temperatures are reported in Table 1. Brisbane is the hottest city on average and also records the lowest variability in average daily temperature. Melbourne is the coldest on average and has a relatively high variability in average daily temperature. Perth has the most variable daily temperatures. There are significant differences in all the cities between the sample means of temperature pre- and post-1950. This suggests that a time trend will be an important component of a model of average daily temperatures.[6] Interestingly, any trend in daily temperatures seems to be driven by the increasing minimum value of daily temperatures rather than by an increasing maximum value.

		Summary Statistics					
	Dates	N	Mean	Med.	S. Dev.	Max.	Min
Brisbane	1887 - 2006	43800	20.52	20.85	4.05	34.65	8.30
Brisbane	1887 - 1949	22995	20.39	20.70	4.11	34.65	8.30
Brisbane	1950 - 2006	20805	20.67	21.00	3.97	34.15	8.45
Melbourne	1856 - 2006	55115	14.95	14.40	4.74	34.55	2.25
Melbourne	1856 - 1949	34310	14.64	14.15	4.72	34.20	2.25
Melbourne	1950 - 2006	20805	15.46	14.90	4.72	34.55	3.80
Perth	1897 - 2006	40150	18.07	17.25	4.94	36.95	6.25
Perth	1897 - 1949	19345	17.92	17.20	4.72	36.95	6.25
Perth	1950 - 2006	20805	18.21	17.25	5.15	36.80	6.25
Sydney	1859 - 2006	54020	17.66	17.80	4.28	33.75	6.40
Sydney	1859 - 1949	33215	17.34	17.50	4.32	33.70	6.40
Sydney	1950 - 2006	20805	18.18	18.25	4.15	33.75	7.70

Table 1: Mean, median, standard deviation, maximum and minimum of average daily temperature in four Australian cities. Note that the sample is curtailed to end on 31 December 2006 to ensure that summary statistics are computed over complete years.

Figure 1 shows the long-term expected values (upper panel) and standard deviations (lower panel) of daily temperatures for each day of the year. Figure 1 shows that all the cities have similar seasonal fluctuation and that the estimates of the long-term expected values of temperature on each day in every city is converging. By contrast, Figure 1 demonstrates more variability in the seasonal pattern of the volatility of temperatures across the cities. It is also noticeable that, despite the length of the temperature records, the estimates of daily volatility appear not to have converged to the same extent as the estimates of the mean temperature.

[6] Given the location of the actual weather stations from which the time-series data are assembled, it is conjectured that this time trend is probably due to urbanisation rather than a manifestation of global warming.

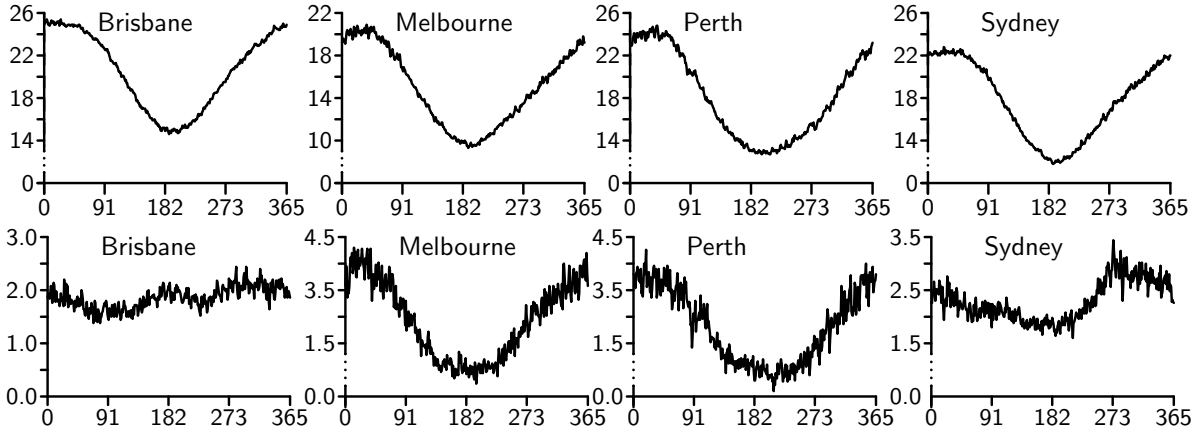

Figure 1: The expected value of the average daily temperatures (upper panel) and the expected value of the volatility of average daily temperatures (lower panel) are shown for Brisbane, Melbourne, Perth and Sydney.

3 Tick Values of Temperature Options

The most commonly referenced weather indices on which temperature derivatives are written are cumulative heating degree days (HDDs) and cumulative cooling degree days (CDDs). Let $T^{\max}$ and $T^{\min}$ be respectively the maximum and minimum temperatures in degrees Celsius measured on a particular day at a specific weather station. The HDD and CDD indices at that station on that day are defined respectively by

$$\begin{aligned} \mathrm{HDD} &= \max\big(0,\, 18 - T\big)\,, \\ \mathrm{CDD} &= \max\big(0,\, T - 18\big)\,, \end{aligned} \tag{1}$$

where T is the arithmetic mean of the maximum and minimum temperatures achieved on that day, namely

$$T = \frac{T^{\max} + T^{\min}}{2}\,. \tag{2}$$

The choice of threshold, in this instance 18°C, is set by market convention and is the standard used in the US. In the southern (northern) hemisphere the HDD (CDD) season would be from May to September, while the CDD (HDD) season would be from November to March.

Temperature-based options are based on *cumulative* heating or cooling degree days constructed by summing daily HDD/CDD indices over a period of N days to get

$$\begin{aligned} \mathrm{H}_N &= \sum_{k=1}^{N} \max\big(0,\, 18 - T_k\big)\,, \\ \mathrm{C}_N &= \sum_{k=1}^{N} \max\big(0,\, T_k - 18\big)\,, \end{aligned} \tag{3}$$

where T_k is the mean temperature, defined as in equation (2), on the kth day of the life of the option. Without loss of generality, the analysis of this paper will be limited to considering

European call options written on cumulative CDDs. The choice of European option is not limiting in the sense that many more complex derivative strategies are in fact combinations of simple European options. The choice of CDDs is more pragmatic, driven by the fact that CDDs are uniformly important to all the major Australian cities in the data set.

Let D be the strike price of a temperature based option defined as a particular value of the relevant cumulative index. The buyer of a vanilla European call option pays an up-front premium and receives a payout if the value of the relevant index exceeds the strike price, D, at the maturity of the option. The tick value of an CDD call option with strike price D and duration N days is therefore

$$\mathcal{T}_N = \max\left(\mathrm{C}_N - D, 0\right). \tag{4}$$

The actual monetary payoff from the contract is the product of the tick value and the tick size, defined as the cash value of a tick. Given the probability density function, $f_N(x)$ of the relevant cumulative index over the period of the contract, a call option for N days with strike price D, for example, will have expected tick-value

$$\mathrm{E}\left[\,\mathcal{T}_N\,\right] = \int_D^\infty (x - D)\, f_N(x)\, dx\,.$$

Traditionally, the valuation of options under schemes such as that of Black and Scholes (1973) discounts the expected payoff at the risk-free force of interest. This choice of discount rate is based on a zero-arbitrage argument involving the formation of a portfolio consisting of a risk-free combination of an option and the underlying asset. However, in context of a temperature-based weather derivative, the underlying indices are not tradable, and therefore these derivatives cannot be priced by means of a zero-arbitrage argument. Therefore the focus turns to estimating the distribution of payoffs for pricing purposes.

The most common practical approach used to price temperature-based derivatives is the *actuarial* valuation method, discussed, for example, in Zeng (2000) and Platen and West (2003). Broadly speaking this approach prices the derivative at the mean expected payoff plus a premium for overhead expenses. The simplest way of implementing this pricing scheme is to review historical records of C_N over the period relevant to the contract and use these values to calculate its hypothetical payoff. The actuarially fair price for the derivative would then be the mean historical payoff.

This approach is only sensible if the values of C_N are independent and identically distributed random variables. Figure 2 illustrates the sequence of historical records of C_N for the period 1 January to 31 March for each of the four cities. A cursory inspection of Figure 2 suggests that a time trend is present in the historical record of cumulative cooling degree days in all of the cities. To test this hypothesis formally, a simple quadratic trend model is proposed. Although

the quadratic term is not expected to be significant, it is included to account for the possibility of piecewise trends in cumulative CDDs due to the effect of urbanisation late in the sample period. Accordingly, cumulative CDDs are described by the general model

$$C_t = \eta_0 + \eta_1 \text{ Trend}_t + \eta_2 \text{ Trend}_t^2 + \epsilon_t$$

where ϵ_t is now distributed $iid(0, \sigma_\epsilon^2)$.

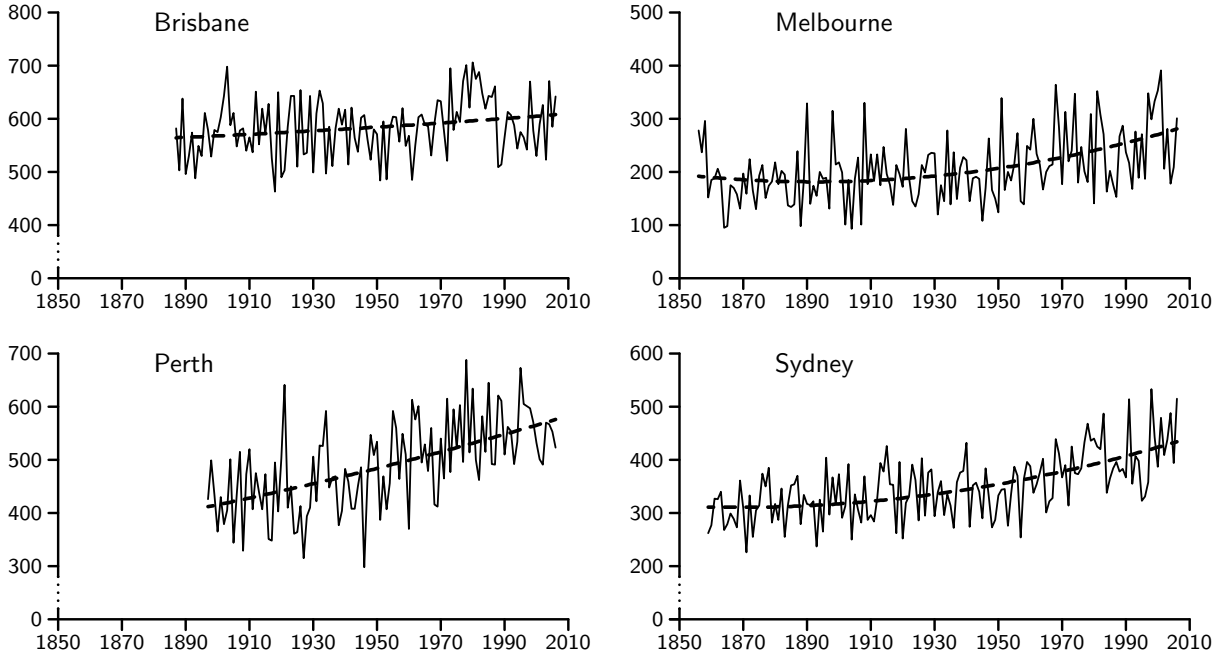

Figure 2: Time series of cumulative CDDs for each city with estimated time trend superimposed (dashed line).

Estimation of the parameters of this model for each city yields

$$\begin{aligned}
\mathrm{E}[C_t]_{\text{Brisbane}} &= \underset{(14.8071)}{564.0290} + \underset{(0.5603)}{0.2617} \text{ Trend}_\text{t} + \underset{(0.0044)}{0.0008} \text{ Trend}_t^2 \\
\mathrm{E}[C_t]_{\text{Melbourne}} &= \underset{(14.1135)}{192.4003} - \underset{(0.4259)}{0.5819} \text{ Trend}_\text{t} + \underset{(0.0027)}{0.0077} \text{ Trend}_t^2 \\
\mathrm{E}[C_t]_{\text{Perth}} &= \underset{(19.7863)}{410.2985} + \underset{(0.8155)}{1.2290} \text{ Trend}_\text{t} + \underset{(0.0071)}{0.0025} \text{ Trend}_t^2 \\
\mathrm{E}[C_t]_{\text{Sydney}} &= \underset{(11.6784)}{311.1655} - \underset{(0.3595)}{0.1276} \text{ Trend}_\text{t} + \underset{(0.0023)}{0.0065} \text{ Trend}_t^2
\end{aligned}$$

where the figures in parentheses are standard errors. There is enough evidence in these results to conclude that the trend in cumulative CDDs is statistically significant which leads inexorably to the conclusion that the cumulative CDDs are not identically distributed and as shown by Clements *et al.* (2008) this fact leads to simple pricing based on historical records being unreliable.

4 A Continuous-time Autoregressive Model of Temperature

4.1 The model

For all cities, the temperature, T_t, is the average daily temperature defined in equation (2). Following the general convention (Davis, 2001, Alaton *et al.*, 2002, Benth and Šaltynė-Benth, 2005), the deviations of temperature from its long-term average $\theta_t = T_t - \bar{T}_t$ are modeled as a low-order autoregressive (AR) process[7]

In this context, the daily average temperature $T(t)$ is expressed in the form $\overline{T}(t)+\theta(t)$ with $\overline{T}(t)$ modelling the mean average temperature at that time and $\theta(t)$ modelling the deviation of the average daily temperature from the seasonal mean temperature. The process $\theta(t)$ is assumed to satisfy the stochastic differential equation

$$d\theta = -\alpha\theta\, dt + \sigma(t)\, dW\,, \tag{5}$$

where dW is the increment in the Wiener process, and the parameter α (assumed constant) and the function $\sigma(t)$ are to be determined from observations of average daily temperature.

To utilise this model for predicting the predicting the payoff from temperature-based derivatives an estimate of the parameter α is required. To do so, it is necessary to obtain estimates of $\overline{T}(t)$ and $\sigma(t)$ in the form of flexible functions that can accommodate their anticipated seasonal behaviour. Fourier series of low order therefore provide an ideal specification for $\overline{T}(t)$ and $\sigma(t)$, which henceforth will be assumed to be represented by the respective generic forms

$$\left.\begin{aligned} \overline{T}(s) &= a_0 + b_0 s + \sum_{k=1}^{n} a_k \cos(\omega_k s) + b_k \sin(\omega_k s)\,, \\ \sigma^2(s) &= c_0 + \sum_{k=1}^{n} c_k \cos(\omega_k s) + d_k \sin(\omega_k s)\,, \end{aligned}\right] \quad \omega_k = \frac{2k\pi}{365}\,, \tag{6}$$

where $s = 0$ is assumed to be the calender date of the first observation of average daily temperature. The contribution $b_0 s$ in the expression for $\overline{T}(s)$ is present to take account of any annual trend in daily average temperature. Otherwise expressions (6) assume that seasonal variations in daily average temperature follow an annual cycle which is independent of calendar year. We describe two strategies to determine the value of α and the coefficients in the Fourier series (6).

4.2 Parameter estimation

Suppose that the data consists of observations of daily average temperatures $T_1, T_2, \cdots, T_N$ at the increasing sequence of times $t_1, t_2, \cdots, t_N$. The essence of the regression approach is that

[7] Alternatively, a fractionally integrated process for deviations could be used (see, for example, Caballero and Jewson, 2002), but this modeling avenue is not pursued here.

the Fourier coefficients of $\overline{T}(s)$ can be constructed in a straightforward way by minimising the objective function

$$\Psi(a_0, b_0, a_1, \cdots, b_n) = \sum_{j=1}^{N} \left(T_j - \overline{T}(t_j)\right)^2$$

by suitable choice of the coefficients a_0, $b_0, \cdots, b_n$. Once these coefficients are determined and the expression for $\overline{T}(s)$ is known, then the residuals $\theta_1, \theta_2, \cdots, \theta_n$ can be computed directly from the formula $\theta_j = T_j - \overline{T}(t_j)$, and the problem is now to find the values of α and the coefficients $c_0, c_1, d_1, \cdots, d_n$ which best fit the residuals $\theta_1, \theta_2, \cdots, \theta_n$. One possible way to achieve this objective is to note that equation (5) has solution

$$\theta(t) = \int_{-\infty}^{t} e^{-\alpha(t-s)} \sigma(s)\, dW(s). \tag{7}$$

This solution satisfies $\mathrm{E}\left[\theta(t)\right] = 0$ with autocorrelation function at lag u given by

$$\mathrm{E}\left[\theta(t)\theta(t+u)\right] = e^{-\alpha u} S(t)\,, \qquad S(t) = \int_{-\infty}^{t} e^{-2\alpha(t-s)} \sigma^2(s)\, ds\,, \tag{8}$$

where $S(t)$ denotes the seasonal variance of the deviation of daily average temperature from its mean value. The function $S(t)$ may be estimated directly from the data, and will of course inherit the cyclical behaviour of $\sigma^2(t)$. It is a straightforward to demonstrate that $\sigma^2(t)$ and $S(t)$ are connected by the identity

$$\sigma^2(t) = \frac{dS(t)}{dt} + 2\alpha S(t)\,.$$

Consequently, the expression for $S(t)$ corresponding to the expression (6) for $\sigma^2(s)$ is

$$S(s) = p_0 + \sum_{k=1}^{n} p_k \cos(\omega_k s) + q_k \sin(\omega_k s)\,, \quad \omega_k = \frac{2k\pi}{365}\,, \tag{9}$$

where the parameters $c_0, c_1, \cdots, d_n$ are related to the parameters $p_0, p_1, \cdots, q_n$ by the formulae

$$c_0 = 2\alpha p_0\,, \qquad \left.\begin{aligned} c_k &= 2\alpha p_k + \omega_k q_k\,, \\ d_k &= -\omega_k p_k + 2\alpha q_k\,, \end{aligned}\right] \qquad k = 1, 2. \cdots, n\,. \tag{10}$$

Equations (9) and (10) supply the first part of the algorithm to determine the value of α and the Fourier coefficients in the specification of $\sigma^2(s)$.

The second part of the algorithm is based on a result of Bibby and Sorensen (1995) concerning the properties of the solution of the initial value problem for the stochastic differential equation $dX_t = \alpha(\theta - X_t)\, dt + \sigma(X_t) dW_t$ in which $\sigma(X_t)$ is a positive real-valued function. They show that an unbiased estimate of the parameter α is given by the expression

$$\alpha = -\log\left[\frac{\left(\sum_{k=1}^{n} \frac{X_{k-1}}{\sigma^2(X_{k-1})}\right)\left(\sum_{k=1}^{n} \frac{X_k}{\sigma^2(X_k)}\right) - \left(\sum_{k=1}^{n} \frac{X_{k-1}X_k}{\sigma^2(X_{k-1})}\right)\left(\sum_{k=1}^{n} \frac{1}{\sigma^2(X_{k-1})}\right)}{\left(\sum_{k=1}^{n} \frac{X_{k-1}}{\sigma^2(X_{k-1})}\right)^2 - \left(\sum_{k=1}^{n} \frac{X_{k-1}^2}{\sigma^2(X_{k-1})}\right)\left(\sum_{k=1}^{n} \frac{1}{\sigma^2(X_{k-1})}\right)}\right]. \tag{11}$$

The difficulty, however, in using this expression is that $\sigma^2(X_t)$ is unknown whereas what is known is the seasonal variance of the residuals. The strategy for finding the values of α and the coefficients $c_0, \cdots, d_n$ is therefore the following.

First compute the Fourier coefficients of $S(t)$ directly from the deviations $\theta_1, \theta_2, \cdots, \theta_N$ formed from $T_t - \bar{T}_t$. Now choose an arbitrary value for α, say α_0, and compute the Fourier coefficients of $\sigma^2(s)$ from formulae (10) with $\alpha = \alpha_0$. Knowing the Fourier coefficients of $\sigma^2(s)$ enables $\sigma^2(t_k)$ to be computed from the formula (6). Expression (11) is now used to estimate α_1, but as might be anticipated, its value will not be α_0 simply because α_0 was chosen arbitrarily. This procedure defines the first iteration of an algorithm to find the value of α. The procedure can be repeated by recomputing $\sigma^2(t_k)$ taking account of the new value of α and recalculating another value of α from expression (11). This procedure is repeated until consecutive values of α are not deemed to be significantly different, and the coefficients of the Fourier representation of $\sigma^2(s)$ are finally determined from the Fourier representation of the seasonal variance $S(t)$ via formulae (10).

The values of α and the coefficients $a_0, \cdots, b_n$ and $c_0, \cdots, d_n$ can either be used as they stand or can be used as an initial guess for the parameters of the likelihood approach for estimating the values of these parameters outlined in the next subsection.

4.3 Maximum-likelihood estimation

The feasibility of parameter estimation by maximum likelihood (ML) in this instance relies on the fact that the transitional probability density function of average daily temperature can be computed under the assumption that the deviations of average daily temperature from its mean value satisfies the stochastic differential equation (5). Ito's lemma applied to the stochastic differential equation (5) may be shown to lead to the formal solution

$$\theta(t) = \theta_j \, e^{-\alpha(t-t_j)} + \int_{t_j}^{t} e^{-\alpha(t-s)} \sigma(s) \, dW_s \,, \qquad t > t_j \,. \tag{12}$$

with $\theta_j = \theta(t_j)$. The important observation from this solution is that $\theta(t)$ is a Gaussian random variable with mean value $\mathrm{E}\left[\theta(t)\right] = \theta_j \, e^{-\alpha(t-t_j)}$ and variance

$$\chi(t, t_j) = \int_{t_j}^{t} e^{-2\alpha(t-s)} \sigma^2(s) \, ds = S(t) - e^{-2\alpha(t-t_j)} S(t_j) \,, \tag{13}$$

where the latter expression for $\chi(t, t_j)$t is derived directly from the definition of $S(t)$ given in equation (8). Because $T = \overline{T}(t) + \theta(t)$, then the average daily temperature T is itself Gaussian distributed with mean value $\overline{T}(t) + \left(T_j - \overline{T}_j\right) e^{-\alpha(t-t_j)}$ and variance $\chi(t, t_j) = S(t) - e^{-2\alpha(t-t_j)} S(t_j)$ where

$$\overline{T}(t) = a_0 + b_0 t + \sum_{k=1}^{n} a_k \cos(\omega_k t) + b_k \sin(\omega_k t) \,, \qquad \omega_k = \frac{2k\pi}{365} \,. \tag{14}$$

Thus the average daily temperature $T(t)$ has transitional probability density function

$$f(T,t\,|\,T_j,t_j) = \frac{1}{\sqrt{2\pi\chi(t,t_j)}}\,\exp\Big[-\frac{\big(T-\overline{T}(t)-(T_j-\overline{T}_j)\,e^{-\alpha(t-t_j)}\big)^2}{2\chi(t,t_j)}\Big]\,. \tag{15}$$

The likelihood of observing the sequence $T_1, T_2, \cdots, T_N$ of average daily temperatures at calendar times $t_1, t_2, \cdots, t_N$ is therefore

$$\mathcal{L}(\alpha; a_0, \cdots, b_n; c_0, \cdots, d_n) = \prod_{j=1}^{N-1} f(T_{j+1}, t_{j+1}\,|\,T_j, t_j)\,. \tag{16}$$

In practice this likelihood is maximised with respect to the set of parameters $\alpha; a_0, \cdots, b_n; c_0, \cdots, d_n$ by minimising the negative log-likelihood function $-\log\mathcal{L}(\alpha; a_0, \cdots, b_n; c_0, \cdots, d_n)$ which in this instance takes the convenient form

$$\begin{aligned} -\log\mathcal{L} \;=\; & \frac{N-1}{2}\log 2\pi + \frac{1}{2}\sum_{j=1}^{N-1}\log\big(S_{j+1}-e^{-2\alpha(t_{j+1}-t_j)}S_j\big) \\ & +\frac{1}{2}\sum_{j=1}^{N-1}\frac{\big(T_{j+1}-\overline{T}_{j+1}-(T_j-\overline{T}_j)\,e^{-\alpha(t_{j+1}-t_j)}\big)^2}{S_{j+1}-e^{-2\alpha(t_{j+1}-t_j)}S_j}\,, \end{aligned} \tag{17}$$

where the notation $S_j = S(t_j)$ has been used. The optimal values for the parameters of this model are taken to be those which minimise expression (16). Although model (5) is specified in terms of the intrinsic function $\sigma(t)$, from a purely technical point of view it is easier to treat the Fourier coefficients of $S(t)$ as the parameters to be determined by the ML procedure. Furthermore, the numerical effort required to minimise $-\log\mathcal{L}$ can be significantly reduced by taking as starting values the optimal values identified by the regression procedure described in above.

5 Analytical Results and Distributions

There are two distinct factors contributing to the value of temperature-based indices. The first is the stochastic behaviour of the time series of average daily temperatures, and the second is the choice of cut-off temperature above which accumulation of the relevant temperature-based index takes place. To appreciate how the time course of average daily temperature is driven by the transitional probability density function of average daily temperature consider the following argument.

The average daily temperature on the first day of an option, say T_1, is simply a random draw from the marginal density of average daily temperatures at that time of year, namely $T_1 \sim \mathrm{N}\,(\overline{T}_1, S_1)$ where $\overline{T}_1$ and S_1 denote respectively the mean average daily temperature and the variance of average daily temperature at that calendar date. Thereafter, average daily temperatures on consecutive days of the option are correlated through the transitional probability density

function (15). For example, the conditional probability density function of T_2, the average daily temperature on the second day of the option, is

$$\begin{aligned} f_2(T,t_2) = \int_{-\infty}^{\infty} f_1(T_1,t_1) f(T,t_2\,|\,T_1,t_1)\,dT_1 = \frac{1}{\sqrt{2\pi S_1}}\,\frac{1}{\sqrt{2\pi(S_2 - e^{-2(t_2-t_1)}S_1)}} \\ \times \int_{-\infty}^{\infty} \exp\Big[-\frac{\big(T_1-\overline{T}_1\big)^2}{2S_1}\Big] \exp\Big[-\frac{\big(T-\overline{T}(t)-(T_1-\overline{T}_1)\,e^{-\alpha(t_2-t_1)}\big)^2}{2(S_2-e^{-2(t_2-t_1)}S_1)}\Big]\,dT_1\,. \end{aligned} \tag{18}$$

Computation of the integral on the right hand side of equation (18) leads to the result that $T_2 \sim \mathrm{N}\,(\overline{T}_2, S_2)$ where $\overline{T}_2$ and S_2 denote respectively the mean average daily temperature and the variance of average daily temperature on the second day of the option. In other words, the value of T_2 is a draw from the marginal distribution of average daily temperature on the calendar date corresponding to the second day of the option. This argument may be continued to each day in the life of the option. The conclusion is that the average daily temperature on any day of the option is a random draw from the marginal density of average daily temperature for the calendar date corresponding to that day.

Although this result has been demonstrated explicitly in this instance for the nonhomogeneous Ornstein Uhlenbeck process, logically this is a generic result. In the absence of explicit values of average daily temperature during the lifetime of an option written on a temperature index, the best estimate of average daily temperature on a day in the lifetime of the option is the marginal density of average daily temperature on the calender date corresponding to that day. Because temperature-based weather derivatives are traded well before any temperature information becomes available for the period of the option, the first step in pricing any temperature-based option is therefore to recognise that the expected value of the temperature index on which that option is written will be determined by the marginal density of average daily temperature on the calendar dates corresponding to each day in the lifetime of the option. The second step in the pricing strategy is to realise that the daily contributions made to the temperature index on which a temperature-based option is written inherit a correlation structure since the average daily temperature is itself correlated in time. The successful pricing of temperature-based options relies crucially on the extent to which the effect of this correlation in the temperature index can be quantified.

An unwelcome complication is this challenge is the issue that it cannot be assumed *a priori* that each day in the lifetime of a temperature-based option will make a non-zero contribution to the temperature index on which that option is written. A special case and the general case are now considered in detail.

5.1 Special case

Consider the special situation in which every day of the option delivers a non-zero contribution to the temperature index with very high probability, assumed in this analysis to be unity. The cumulative tick value of a CDD option starting at calendar date t_j and of duration m days is therefore

$$\mathcal{T}_{j,m} = \int_{t_j}^{t_{j+m}} \big(T(t) - 18\big)\, dt = \int_{t_n}^{t_{j+m}} \big(\overline{T}(t) + \theta(t) - 18\big)\, dt\,. \tag{19}$$

Because $\mathrm{E}\,[\theta(t)] = 0$, the expected value of the payoff from such an option is

$$\mathrm{E}\,[\mathcal{T}_{j,m}] = \int_{t_j}^{t_{j+m}} \big(\overline{T}(t) - 18\big)\, dt\,, \tag{20}$$

and the variance of this payoff is

$$\begin{aligned} \mathrm{E}\,\big[\big(\mathcal{T}_{j,m} - \mathrm{E}\,[\mathcal{T}_{j,m}]\big)^2\big] &= \mathrm{E}\,\Big[\Big(\int_{t_j}^{t_{j+m}} \theta(u)\, du\Big)\Big(\int_{t_j}^{t_{j+m}} \theta(v)\, dv\Big)\Big] \\ &= \int_{t_j}^{t_{j+m}} \int_{t_j}^{t_{j+m}} \mathrm{E}\,[\theta(v)\theta(u)]\, du\, dv\,. \end{aligned} \tag{21}$$

The computation of this integral begins by dividing the region of integration into the regions above and below the line $u = v$ and noting that

$$\mathrm{E}\,[\theta(u)\theta(v)] = \left[\begin{array}{ll} e^{-\alpha(u-v)}\, S(v) & u \geq v\,, \\ e^{-\alpha(v-u)}\, S(u) & v \geq u\,, \end{array}\right.$$

it follows directly that

$$\begin{aligned} \mathrm{E}\,\big[\big(\mathcal{T}_{j,m} - \mathrm{E}\,[\mathcal{T}_{j,m}]\big)^2(t)\big] &= 2\int_{t_j}^{t_{j+m}} \Big(\int_{u}^{t_{j+m}} e^{-\alpha(v-u)}\, S(u)\, dv\Big)\, du \\ &= \frac{2}{\alpha}\int_{t_j}^{t_{j+m}} S(u)\big[1 - e^{-\alpha(t_{j+m}-u)}\big]\, du\,. \end{aligned} \tag{22}$$

Direct calculation based on the ansatz (6) for $\overline{T}(t)$ and expression (20) leads to the formula

$$\mathrm{E}\,[\mathcal{T}_{j,m}] = m(a_0 - 18) + m b_0 t_{j+m/2} + 2\sum_{k=1}^{n} \sin\phi_k \frac{a_k \cos(\omega_k t_j + \phi_k) + b_k \sin(\omega_k t_j + \phi_k)}{\omega_k} \tag{23}$$

for the mean of the cumulative tick value for average daily temperature. Furthermore, the variance of this cumulative tick value is given by expression (22) and leads to the formula

$$\begin{aligned} \mathrm{E}\,\big[\big(\mathcal{T}_{j,m} - \mathrm{E}\,[\mathcal{T}_{j,m}]\big)^2\big] &= \sum_{k=1}^{n} p_k \Big[\frac{\sin\omega_k t_{j+m} - \sin\omega_k t_j}{\omega_k} - I_{j,m,k}\Big] \\ &\quad + \sum_{k=1}^{n} q_k \Big[\frac{\cos\omega_k t_j - \cos\omega_k t_{j+m}}{\omega_k} - J_{j,m,k}\Big] + \frac{2p_0}{\alpha^2}\big(m\alpha - 1 + e^{-m\alpha}\big) \end{aligned} \tag{24}$$

where $I_{j,m,k}$ and $J_{j,m,k}$ denote the values of the integrals

$$\begin{aligned} I_{j,m,k} &= \int_{t_j}^{t_{j+m}} \big[1 - e^{-\alpha(t_{j+m}-u)}\big] \cos(\omega_k u)\, du\,, \\ J_{j,m,k} &= \int_{t_j}^{t_{j+m}} \big[1 - e^{-\alpha(t_{j+m}-u)}\big] \sin(\omega_k u)\, du\,. \end{aligned}$$

5.2 General case

Consider now a CDD call option of duration m days starting at calendar date t_j. The k-th day in the lifetime of this option will contribute to the temperature index driving the value of the option with probability

$$p_k = \Phi(z_k)\,, \qquad z_k = \frac{\overline{T}_{j+k} - 18}{\sqrt{S_{j+k}}}\,, \tag{25}$$

where $\Phi(z)$ is the cumulative distribution function of the standard normal. The cumulative tick value of a CDD option of duration m days starting at calendar date t_j is

$$C_{j,\,m} = \sum_{k=1}^{m} \mathcal{T}_{j+k}\,, \qquad \mathcal{T}_{j+k} = \max[T_{j+k} - 18, 0]\,. \tag{26}$$

The expected value and variance of $C_{j,\,m}$ are respectively

$$\begin{aligned} \mathrm{E}\left[C_{j,\,m}\right] &= \sum_{k=1}^{m} \mathrm{E}\left[\mathcal{T}_{j+k}\right] \\ \mathrm{Var}\left[C_{j,\,m}\right] &= \sum_{k=1}^{m} \mathrm{E}\left[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})^2\right] + 2\sum_{k=1}^{m-1}\sum_{r=k+1}^{m} \mathrm{E}\left[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})(\mathcal{T}_{j+r} - \overline{\mathcal{T}}_{j+r})\right]. \end{aligned} \tag{27}$$

Each of these expressions is considered in turn.

5.2.1 Mean value of the temperature index of CDDs

Because T_{j+k} is a Gaussian random variable with mean value $\overline{T}_{j+k}$ and variance S_{j+k}, then the expected value of the temperature index for a CDD call option of duration m days starting at calendar date t_j is

$$\mathrm{E}\left[\mathcal{T}_{j+k}\right] = \frac{1}{\sqrt{2\pi S_{j+k}}} \int_{18}^{\infty} (T - 18) \exp\Big[- \frac{(T - \overline{T}_{j+k})^2}{2S_{j+k}} \Big]\, dT$$

The use of the change of variable $T = \overline{T}_{j+k} + z\sqrt{S_{j+k}}$ in this integral gives

$$\mathrm{E}\left[\mathcal{T}_{j+k}\right] = \frac{\sqrt{S_{j+k}}}{\sqrt{2\pi}} \int_{z_{j+k}}^{\infty} (z - z_{j+k})\, e^{-z^2/2}\, dz\,, \qquad z_{j+k} = \frac{18 - \overline{T}_{j+k}}{\sqrt{S_{j+k}}}$$

which in turn can be expressed in terms of $\phi(z)$, the probability density function of the standard normal, and $\Phi(z)$, the cumulative distribution function of the standard normal. The result of this straightforward calculation is that

$$\mathrm{E}\left[\mathcal{T}_{j+k}\right] = \sqrt{S_{j+k}}\left[\phi(z_{j+k}) + z_{j+k}\Phi(z_{j+k})\right], \qquad z_{j+k} = \frac{18 - \overline{T}_{j+k}}{\sqrt{S_{j+k}}}\,, \tag{28}$$

which in turn leads to the general result

$$\mathrm{E}\left[C_{j,\,m}\right] = \sum_{k=1}^{m} \sqrt{S_{j+k}}\left[\phi(z_{j+k}) + z_{j+k}\Phi(z_{j+k})\right]. \tag{29}$$

5.2.2 Variance of the temperature index of CDDs

The primary difficulty in computing the variance of the temperature index lies in the fact that the daily contributions to this index are correlated random variables thereby making $C_{j,m}$ a sum of correlated random variable with point density at zero. The analysis considered here treats separately the contributions from the first and second terms on the right hand side of equation (27).

The key idea in constructing this variance is to imagine the sample space of realisations of average daily temperature over the interval $[t_j, t_{j+m}]$, and consider the behaviour of a particular day during this period. If calendar day t_{j+k} always makes a nonzero contribution to the value of temperature index then the variance of this contribution is S_{j+k}, the variance of θ_{j+k} on that day. On the other extreme, if this day never contributes to the value of the temperature index then the variance of its contribution is zero. Therefore if calendar day t_{j+k} contributes to the value of the temperature index on fraction p_{j+k} of days then an interpolation argument suggests that $p_{j+k}S_{j+k}$ is a reasonable estimate for the value of $\mathrm{E}\,[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})^2]$. Based on this idea, the first summation on the right hand side of equation (27) has approximate values

$$\sum_{k=1}^{m} \mathrm{E}\,[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})^2] \approx \sum_{k=1}^{m} p_{j+k}S_{j+k}\,. \tag{30}$$

The second summation on the right hand side of equation (27) is a correction to expression (30) reflecting the fact that contributions to the value of the temperature index from different days are not independent. The contribution made by the quantity $\mathrm{E}\,[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})(\mathcal{T}_{j+r} - \overline{\mathcal{T}}_{j+r})]$ to the variance of the temperature index is argued in a similar way. In the absence of clipping, the variance of this product is equal to $\mathrm{E}\,[\theta_{j+k}\theta_{j+r}]$ with value $S_{j+k}\,e^{-\alpha(r-k)}$ assuming that $r > k$. However, the product $\mathcal{T}_{j+k}\mathcal{T}_{j+r}$ is nonzero with probability $p_{j+k}p_{j+r}$ and therefore the same interpolation argument indicates that $\mathrm{E}\,[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})(\mathcal{T}_{j+r} - \overline{\mathcal{T}}_{j+r})]$ is reasonably estimated by $p_{j+k}p_{j+r}S_{j+k}\,e^{-\alpha(r-k)}$. Based on this idea, the second summation on the right hand side of equation (27) has approximate values

$$2\sum_{k=1}^{m-1}\sum_{r=k+1}^{m} \mathrm{E}\,[(\mathcal{T}_{j+k} - \overline{\mathcal{T}}_{j+k})(\mathcal{T}_{j+r} - \overline{\mathcal{T}}_{j+r})] \approx 2\sum_{k=1}^{m-1} p_{j+k}S_{j+k} \sum_{r=k+1}^{m} p_{j+r}\,e^{-\alpha(r-k)}\,. \tag{31}$$

In conclusion, the variance of $C_{j,m}$ is well approximated by the formula

$$\mathrm{Var}\,[C_{j,m}] = \sum_{k=1}^{m} p_{j+k}S_{j+k} + 2\sum_{k=1}^{m-1} p_{j+k}S_{j+k} \sum_{r=k+1}^{m} p_{j+r}\,e^{-\alpha(r-k)}\,. \tag{32}$$

5.2.3 Evaluation of variance

The variance of $C_{j,m}$ can be computed efficiently by means of the auxiliary function

$$\psi_{j,k} = \left[\begin{array}{ll} \displaystyle\sum_{r=k+1}^{m} p_{j+r}\, e^{-\alpha(r-k)} & k < m \\ 0 & k = m \end{array} \right. \tag{33}$$

using the formula

$$\mathrm{Var}\,[C_{j,m}] = \sum_{k=1}^{m} p_{j+k} S_{j+k} \big(1 + 2\psi_k\big)\,. \tag{34}$$

However, rather than computing $\psi_{j,k}$ by evaluating each individual sum, it is convenient to evaluate each function using the iterative property

$$\begin{aligned} \psi_{j,m} &= 0\,, \\ \psi_{j,k-1} &= e^{-\alpha}(\psi_{j,k} + p_{j+k})\,, \quad k \le m\,. \end{aligned} \tag{35}$$

To appreciate the equivalence of definition (33) and property (35), note that for $k < m$ the function ψ_k satisfies

$$\begin{aligned} \psi_{j,k-1} - e^{-\alpha}\psi_{j,k} &= \sum_{r=k}^{m} p_{j+r} e^{-\alpha(r-k+1)} - e^{-\alpha} \sum_{r=k+1}^{m} p_{j+r} e^{-\alpha(r-k)} \\ &= p_{j+k} e^{-\alpha} + \sum_{r=k+1}^{m} p_{j+r} e^{-\alpha(r-k+1)} - \sum_{r=k+1}^{m} p_{j+r} e^{-\alpha(r-k+1)} = p_{j+k} e^{-\alpha}\,. \end{aligned}$$

6 Computing Expected Payoffs

The task is now to provide a means of gauging the performance of the method suggested in Section 5 for computing the expected payoffs of contracts. In doing so, the performance of the proposed method is compared to the historical approach of Zeng (2000) and Platen and West (2003). In this paper, the metric for comparison is taken to be the mean 'profit' of a 90-day call option contract taken over a period of years. Profit is defined from the point of view of the buyer of the call option as the difference between the actual tick value of the contract and the expected tick value or 'price' of the option. Of course, this is not meant to represent a true price for the option, as this notional pricing strategy takes no account of discounting or overhead expenses. But of course, any pricing scheme will stand or fall by its ability to estimate the expected tick value accurately.

The descriptive statistics of the cumulative CDDs upon which historical pricing is based are reported in Table 2. These are very much as expected given the geographical locations of the cities, but there are, however, two observations of note arising out of Table 2. It is apparent that the distribution of cumulative CDDs for Melbourne is skewed to the right as evidenced by

a mean which is significantly larger than the median. This is to be expected given both the instances of extreme heat in Melbourne and the strength of the trend in the Melbourne CDD data identified in Section 3. Perth, on the other hand, is notable for the diffuse nature of the distribution of cumulative CDDs, recording a standard deviation significantly larger than those of the other cities.

	Summary Statistics					
	N	Mean	Med.	S. Dev.	Min.	Max.
Brisbane	121	584.2	584.6	54.49	463.3	705.9
Melbourne	152	207.9	195.6	64.09	93.5	391.4
Perth	111	489.6	492.2	83.30	298.3	688.3
Sydney	149	350.0	350.2	60.07	225.5	533.3

Table 2: Mean, median, standard deviation, minimum and maximum cumulative CDDs in four Australian cities.

It is also instructive to examine the distributions of cumulative CDDs in each of the cities considered. Figure 3 plots both the distribution of historical cumulative CDDs (shaded region) and the predicted distributions for 1950 (dashed line) and 2007 (solid line) generated by closed-form approximations to the distributions of CDDs derived here. To the uniformed eye, the distribution of historical cumulative CDDs may appear well behaved and taken as reasonable evidence in favour of using historical records to price temperature-based derivatives. When compared to the distributions for 1950 and 2007 generated by the analytical approach, the potential for error inherent in the historical approach becomes evident. Not only does the mean of the predicted distribution change noticeably over time, as would be expected given the discussion in Section 3, but the distribution also has lower volatility.

The profits generated by two call-option contracts with different strike prices, written on the period 1 January to 31 March are now reported in Tables 3 and 4 respectively. The experiments begin by pricing these options for the year 1950 using data up to and including 1949. The actual payoff for 1950 is recorded, the profit or loss stored and the data set is updated to include all the temperature records for the next year. These steps are repeated up to and including 2007 giving a total of 58 separate profits for each option. The call options used in the experiment have respective strike prices set to be approximately $D = \mu + 0.5\sigma$ and $D = \mu + 0.75\sigma$ where μ is the unconditional mean and σ is the unconditional standard deviation of CDDs up to the current year under consideration. The means and standard deviations of the profits are regarded

as measures of the performance of each of the methods used to determine expected tick values.

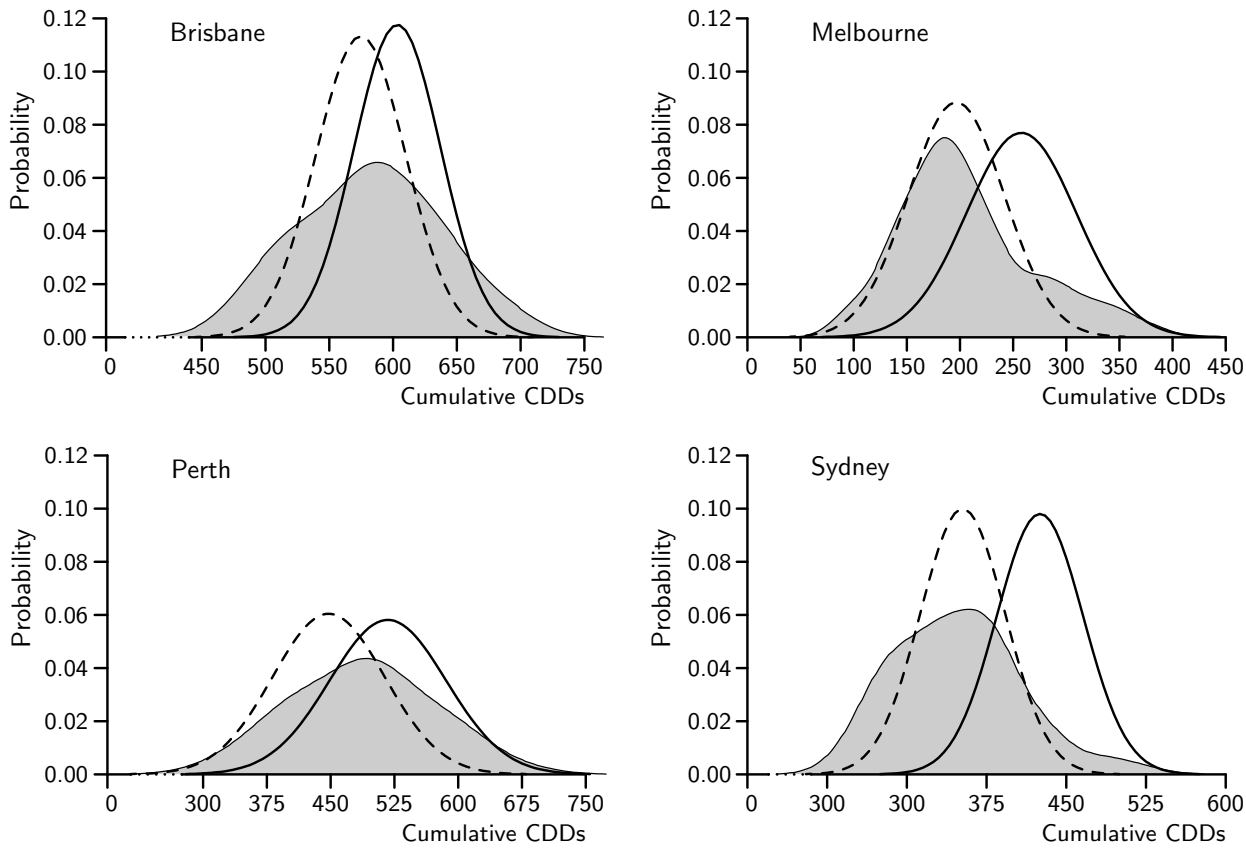

Figure 3: Density of historical cumulative CDDs based on data up to and including 1949 (shaded area), predicted density of cumulative CDD for 1950 (dashed line) and predicted density of cumulative CDD for 2007 (solid line).

The historical pricing reported in Tables 3 and 4 is self-explanatory, but the implementation of the closed-form approximations needs further elucidation. Two variations of this method are implemented, namely an annual version and a quarterly version. The annual approach fits the mean and seasonal variance of average daily temperature using data for the entire year and the best estimates of the parameters are used in the estimation of the relevant distributions. By contrast, the quarterly version focusses on the period from 1 January to the 31 March in each year and fits the mean and seasonal variance of average daily temperature for this period only. In general, the fitting procedure in this interpretation will be implemented only on the period over which the contract is written. The main reason for adopting this approach is that the behaviour of temperature in parts of the year unrelated to the period of the option are not being allowed to influence parameter estimates for the mean and variance processes. Another benefit of this approach is that better resolution of the mean and variance processes with the same number of parameters.

	Brisbane D = 600	Melbourne D = 240	Perth D = 530	Sydney D = 380
Historical				
Mean Payoff	−8.10	−14.31	−23.79	7.84
SDev Payoff	33.11	45.79	43.23	48.88
Quarterly Model				
Mean Payoff	7.17	13.22	2.16	11.66
SDev Payoff	29.64	41.46	41.83	35.54
Annual Model				
Mean Payoff	5.78	15.42	18.30	4.02
SDev Payoff	29.11	41.36	40.04	34.59

Table 3: Means and standard deviations of profits to a 90-day call option defined on CDDs with strike price D approximately equal to $\mu + 0.5\sigma$, where μ and σ are the unconditional mean and standard deviation of available historical CDDs. The option is priced for each year from 1950 to 2007 inclusive.

	Brisbane D = 620	Melbourne D = 260	Perth D = 550	Sydney D = 400
Historical Model				
Mean Payoff	−17.71	−24.68	−35.11	−4.15
SDev Payoff	25.33	38.32	36.09	42.70
Quarterly Model				
Mean Payoff	6.20	11.88	1.29	9.78
SDev Payoff	22.67	34.16	34.22	30.12
Annual Model				
Mean Payoff	5.49	13.32	13.38	4.56
SDev Payoff	22.40	34.15	36.64	29.22

Table 4: Means and standard deviations of profits to a 90-day call option defined on CDDs with strike price D approximately equal to $\mu + 0.75\sigma$, where μ and σ are the unconditional mean and standard deviation of available historical CDDs. The option is priced for each year from 1950 to 2007 inclusive.

The first striking conclusion to be drawn from these results is just how bad historical pricing performs for the Australian temperature data. Interestingly enough, it appears that historical pricing in three of the cities has substantially over-priced the call options. This result is counter-intuitive as it is be expected that the presence of a significant trend in the cumulative CDDs identified in Section 3 would cause the options to be under-priced. The resolution of this conundrum is to be found in the behaviour of temperature between the years 1890 and 1920. In Brisbane, Melbourne and Perth substantial outliers in cumulative CDDs were recorded the likes of which were not seen again until late in the sample period. These outliers would have a disproportionate affect on the pricing of temperature derivatives in the 1960s, 1970s and 1980s. The existence of these outliers would also explain the deterioration in the profit when moving from the from the lower to the higher exercise price when using historical pricing. The weather station in Sydney where the temperature data were recorded did not show these extreme temperature events and consequently historical pricing for Sydney performs significantly better than for the other cities.

Taken as a whole, the closed-approximations used to price the call options generate mean profits closer to zero and with lower standard deviations than historical pricing. Nevertheless, this method appears to underprice somewhat, even though these pricing errors are smaller in magnitude than those generated by the historical method. This underpricing is again a manifestation of the outliers in cumulative CDDs but in this case, not enough weight is given to them. There is little difference in terms of performance of quarterly and annual models, with the exception of Perth where the quarterly model performs better. It is conjectured that this is due to the ability of the quarterly model to better resolve the extreme temperature variations that are prone to take place in Perth. Unlike the case documented for historical pricing, there seems little difference in performance when moving from the lower to the higher exercise price for the the closed-form approach. Overall, it seems as though the analytical method is superior, with these differences being due to differences in the distributions upon which pricing is based, recall earlier discussion of Figure 3.

7 Conclusion

This paper has derived closed-form expressions for approximating the distribution of temperature indices. The major practical use for these approximations is in estimating the payoffs to temperature-based weather derivatives. Although the cumulative cooling degree day index is the focus of this research, the methods used are equally applicable to derivatives based on cumulative heating degree days. Common practice when modelling average daily temperature is to regard the deviations of temperature from its expected value as an Ornstein-Uhlenbeck process.

The key result derived in this paper, is that if this model of temperature is adopted, then the distribution of cumulative cooling degree days may be constructed as the sum of truncated, correlated Gaussian deviates. The mean and variance of the resultant Gaussian distribution depend on the parameters of the underlying temperature process and its autocorrelation structure.

The efficacy of these approximate distributions is tested by estimating the payoffs to temperature-based derivatives. Time series data spanning over a hundred years of average daily temperatures in four major Australian cities are used to estimate the payoffs to European call options written on cooling degree days. The robust conclusion to emerge from this line of research is that the closed-form distributions perform more reliably than the historical pricing method that is commonly advocated in the literature.

References

Alaton, P., Djehiche, B. and Stillberger, D. (2002). On Modelling and Pricing Weather Derivatives, *Applied Mathematical Finance*, **9**, 1-20.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.

Benth, F.E. and Šaltynė-Benth, J. (2005). Stochastic Modelling of Temperature Variations with a View Toward Weather Derivatives, *Applied Mathematical Finance*, **12**, 53-85.

Bibby, B.M. and Sorensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*. **1**, 017–039.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637-659.

Caballero, R., Jewson S. and Brix, A. (2002). Long Memory in Surface Air Temperature Detection, Modelin and Application to Weather Derivative Valuation, *Climate Reasearch*, **21**, 127-140.

Campbell and Diebold. (2004). Weather Forecasting for Weather Derivatives. *Journal of the American Statistical Society*, **100**, 6-16.

Clements, A.E., Hurn, A.S. and Lindsay, K. (2008). Estimating the Payoffs of Temperature-based Weather Derivatives. NCER working Papaer no. 33.

Davis M.H.A. (2001). Pricing Weather Derivatives by Marginal Value, *Quantitative Finance*, **1**, 1-4.

Engle, R.F. and Lee, G.G.J. (1993). A Permanent and Transitory Component Model of Stock Return Volatility, University of California, San Diego, Discussion Paper 92-44R, November 1993.

Garman, M., Blanco, C. and Erickson, R. (2000). Weather Derivatives: Instruments and Pricing Issues. *Environmental Finance*, March.

Jewson S. and Caballero, R. (2003). The Use of Weather Forecasts in the Pricing of Weather Dervivatives, *Meterological Applications*, **10**, 377-389.

Platen, E. and West, J. (2003). Fair Pricing of Weather Derivatives, Quantitative Finance Research Centre, University of Technology, Research Paper Series, 106.

Tindall, J. (2006). Weather Derivatives: Pricing and Risk Management Applications. *Unpublished manuscript*, Institute of Actuaries of Australia.

Wilks, D.S. (1995). Statistical Methods in the Atmospheric Sciences. New York: Academic Press.

Zeng, L. (2000). Pricing Weather Derivatives, *Journal of Risk Finance*, **Spring**, 72-78.