Hot or Cold? : A Comparison Of Different Approaches To The Pricing Of Weather Derivatives

Teddy Oetomo

The University of Sydney

Max Stevenson

The University of Sydney

Corresponding Author: Max Stevenson, H69, Finance Discipline, School of Business, Faculty of Economics and Business, The University of Sydney, NSW, Australia 2006. Tel: + 61 2 9036 9103. E-mail: m.stevenson@econ.usyd.edu.au

Abstract

This article reviews six different temperature forecasting models proposed by the prior literature for pricing weather derivatives. Simulation of these models is used to estimate daily temperature and, as a consequence, the metrics used for pricing temperature derivatives. The models that rely on an Autoregressive Moving Average (ARMA) process exhibit a better goodness-of-fit than those that are established under Monte Carlo simulations. However, the superiority of ARMA-type models is not reflected over the forecast horizon. Over that period, the models which rely on Monte Carlo simulations exhibit a tendency to over-forecast the monthly-accumulated Heating Degree Day (AccHDD) index and to under-forecast the monthly-accumulated Cooling Degree Day (AccCDD) index. Alternatively, models established under the ARMA approach both under-forecast and over-forecast the monthly accumulated indices. All models consistently over-forecast the average daily temperature. The most appropriate pricing model varies between cities and months. Finally, the models examined in this study generate a more accurate AccHDD futures price than the price traded on the market. However, the ability of these models to estimate the AccCDD futures price is significantly poorer than that of the market.

Keywords: Weather derivatives, Heating Degree Days, Cooling Degree Days, forecast error, settlement level, spot price, mis-pricing.

JEL Classification: C52, G13, Q40

Hot or Cold? : A Comparison Of Different Approaches To The Pricing of Weather Derivatives

Recently, global financial markets have witnessed a rapid escalation in the popularity of derivatives with non-tradable underlying assets.[1] By definition, the absence of a tradable underlying asset violates the no-arbitrage and market completeness assumptions, core assumptions underlying standard derivative pricing models. As a consequence, the direct application of standard derivative pricing theory to this type of derivative is incorrect.[2] This highlights the need to develop an appropriate model to price these types of assets. Predominantly, prior studies have focused on modelling short-term interest rate behaviour. These studies include those of Brennan and Schwartz (1982), Vasicek (1977), Cox, Ingersoll and Ross (1985) and Chan, Karoly, Longstaff and Sanders (1992).

The recent introduction of trading in weather derivatives has attracted interest in the pricing of derivatives written against another non-tradable underlying asset.[3] Campbell and Diebold (2001) report the rapid growth of the weather derivative market, from US$500 million in 1998 to US$5 billion by 2000. Alaton, Djehiche and Stillberger (2002), Challis (1999), Hanley (1999) and Cao and Wei (2003) attribute such growth to the deregulation of the energy markets and the significant effect that weather risk has on the volatility of the revenue generated by a large number of companies in general, and specifically by generator and retail electricity companies. This study concentrates on the pricing of the most widely traded weather derivative, namely, temperature-based derivative contracts (see Moreno, 2000).

Despite the increase in trading volume, Nicholls (1999) and Hunter (1999) argue that liquidity is still perceived as a crucial issue restricting further growth in the weather derivative market. As pointed out by Hunter (1999) and Cao and Wei (2003), lack of liquidity has contributed to the delay in the development of an accepted pricing model and, as a consequence, has resulted in disagreement on a fair price for the instruments. In turn, this translates into higher volatility and a larger bid-ask spread.[4]

The absence of a universally accepted pricing model for weather derivatives is the result of the absence of an agreed approach to deriving the expected value of the weather-related spot price at maturity. For temperature-based contracts, the spot price level is represented by the monthly-accumulated Heating Degree Day (AccHDD) or the monthly-accumulated Cooling Degree Day (AccCDD) Index. However, the value of the AccHDD or AccCDD Index reverts to zero at the beginning of each month. Thus, in contrast to other financial indices, the compounded value of the spot price at the time a contract is purchased does not represent the expected value of the spot price at maturity. Hence, in pricing AccHDD or AccCDD derivative contracts, the ability to accurately derive the expected value of a particular month's AccHDD or AccCDD Index is crucial. Temperature forecasting models are used to provide interval forecasts of temperature based on simulation and, as a consequence, expected values of the abovementioned temperature-based weather indices.

This study examines six different forecasting models of temperature embedded in the valuation for weather derivatives proposed by prior studies. It is important to recognise that the temperature forecasting models examined in this study do not represent those that are used by the meteorological bureau around the world. Rather, this study limits its analysis to the temperature forecasting models that serve as the underlying model for valuing weather derivatives contracts. By examining six different forecasting models of temperature proposed by prior studies, this paper provides a number of insights. First, during the model estimation (in-sample) period, most models exhibit a tendency to under-forecast the AccHDD and AccCDD indices, as well as the average daily temperature. Second, resorting to Monte Carlo simulations over the forecasting horizon (the out-of-sample period), consistently results in over-forecasting of the AccHDD Index and under-forecasting of the AccCDD Index. On the other hand, the AutoRegressive Moving Average (ARMA) time series modelling approach consistently under-forecasts the AccHDD Index and over-forecasts the

AccCDD Index. Third, of the two approaches, time series models provide a better fit for in-sample periods than do Monte Carlo simulations. Fourth, the forecasting error was generally higher for the CDD months than for the HDD months, signifying greater difficulty in forecasting the CDD than the HDD. This is consistent with Roustant, Laurent, Bay and Carraro (2001), who document a higher price uncertainty for CDD months than for HDD months. A fifth insight is that the accuracy of the proposed forecasting models decreases for forecasts beyond a 30-day, one month-ahead period.

Although controlling for long-term trend in temperature does not markedly improve forecast accuracy, seasonality is documented as a factor influencing forecast quality. Additionally, the best forecasting model varies between months and cities. Also, by comparing the expected prices derived from interval forecasts with the corresponding market prices of AccHDD and AccCDD futures contracts traded on the Chicago Mercantile Exchange (CME), we demonstrate that the expected prices are closer to the settlement levels of the AccHDD contracts than are the corresponding market prices. However, this finding does not hold when expected prices of the settlement level of AccCDD contracts are compared to market prices. Potentially, these findings could result from different levels of risk premiums incorporated by the market participants for the AccHDD and AccCDD contracts. Additionally, as previously mentioned, it is understood that the forecasting models examined in this study are primarily constructed to serve as the underlying models for pricing weather option and, therefore, might not represent the best model for forecasting the settlement level of futures contracts.

This article is outlined as follows. Section I defines the institutional details of the weather derivatives traded on the CME, while Section II provides an overview of prior literature. Section III sets out the methodology employed and Section IV reports the documented results. Section V concludes by discussing directions for future research.

I. Institutional Details

Prior to the introduction in 1997 of weather contracts on the Chicago Mercantile Exchange (CME), all such derivatives were traded on the Over-The-Counter (OTC) market. The CME introduced both futures and options contracts on temperature-based derivatives for ten major cities in the U.S: namely, Atlanta, Chicago, Cincinnati, Dallas, Des Moines, Las Vegas, New York, Philadelphia, Portland and Tucson. Table 1 (Panel A) below reports the CME code for all 10 cities.

<INSERT TABLE 1 ABOUT HERE>

Generally, the value of temperature-based derivatives is derived from the monthly-accumulated HDD (AccHDD) and CDD (AccCDD) Indices. The temperature-based contracts have been noted by Moreno (2000) as the most widely traded weather contracts. A degree day is a measure of the disparity between a day's average temperature and a specified threshold. The daily average temperature, T_t, is defined to be the arithmetic average between the daily's maximum, $T_{max,t}$ and minimum, $T_{min,t}$ temperature recorded between 12:01 A.M. and 12:00 P.M. midnight.[5]

$$T_t = \frac{1}{2}(T_{\max,t} + T_{\min,t}) \qquad (1)$$

Daily HDD and CDD measures the coldness and warmness, respectively, of the daily temperature compared to a threshold level, G:

$$HDD_t = max(0, G - T_t) \qquad (2)$$

$$CDD_t = max(0, T_t - G) \qquad (3)$$

The standard value for G is 65^o F or 18^o C. Alaton, Djehiche and Stillberger (2002) point out that the choice of 65^o F as the industry standard for the threshold level is due to the belief that, for every degree above and below this threshold level, increasing amounts of energy are required for cooling and heating, respectively.

For the most part, HDD and CDD contracts are written on the accumulation of HDD ($X_{HDD,t}$), or CDD ($X_{CDD,t}$) over a defined period which is typically a calendar month.

$$AccHDD_t = \sum_{t=u}^{v} \max(0, G - T_t) \qquad (4)$$

$$AccCDD_t = \sum_{t=u}^{v} \max(0, T_t - G) \qquad (5)$$

The summation limits, *u* and *v*, denote the start and the end of the accumulation period, respectively. Similar to other futures contracts, the monthly-accumulated HDD or monthly-accumulated-accumulated CDD index futures are legally binding agreements to purchase or sell the value of the HDD or CDD index for a specified price and date. Consistent with other traded derivative contracts, the CME Clearing House performs the service of novation. In order to protect itself against assumed risk, the CME Clearing House imposes daily marking-to-market. The full value of the contract is not transferred but rather, a final marking-to-market on the expiry date is based on the level of the cumulative value of the HDD or CDD Index, accumulated over the months of the contract. This method is known as cash settlement. Table 1 (Panel B) outlines the contract specifications for the monthly-accumulated HDD (AccHDD) and CDD (AccCDD) futures contracts.

II. Literature Review

In pricing temperature-based weather derivatives, the ability to determine an interval forecast of daily average temperature is fundamental in deriving the expected value of the spot price at maturity. This can be understood by examining the supply and demand sides of the weather derivatives market.

As pointed out by Campbell and Diebold (2001), the demand side of the weather derivatives market consists of players who are interested in hedging against the unpredictable component of weather fluctuations or, as they put it, “weather noise” (p.3). In order to form appropriate hedging strategies, what is needed is a model (or models) to capture the spatial and temporal characteristics of weather noise. This can be achieved with an appropriate forecasting model for the AccHDD and AccCDD indices over different locations across the market, as well as for various forecasting horizons. The AccHDD and AccCDD indices are indicators of weather noise and have been historically forecasted from temperature forecasts over longer horizons than those usuallly considered for meterological forecasting.

The key question concerning the supply side of this market is how to price weather derivatives. As previously noted, with a non-tradable underlying physical measure (temperature), and an illiquid market in written AccHDD and AccCDD contracts, it is not possible to construct a portfolio of traded assets that replicates the payoff for monthly-accumulated HDD and CDD. By not being able to use a Black-Scholes approach, Davis (2001) notes that pricing is facilitated on an "expected discounted value" basis whereby the underlying (temperature) is used to forecast accumulated weather noise. This is discounted at the risk-free rate. To justify the use of the risk-free rate, Cao and Wei (2003) demonstrate numerically that the market price of risk associated with temperature or, by extension with AccHDD and AccCDD, is insignificant in most cases. The exceptional cases for which it gains importance will be for high levels of risk aversion and random behaviour of the aggregate dividend process. This is unlikely to be the case for a risk-neutral firm with an aggregate dividend process that usually exhibits mean reversion. Further, for the case of a firm, Campbell and Diebold (2001) point out that the generally reliable method of pricing these options by combining temperature forecasts with a utility function to estimate the demand curve for the derivative, reverts to simply using the forecasts. Dropping the utility function corresponds to the assumption of firm risk-neutrality, with a reliance now on the forecasts to estimate the demand function using the joint distribution of future temperature and revenues. The final estimate of the settlement price is the dollar value of the estimate of the AccHDD or the AccCDD, based on the conditional forecast made on the purchase date of the contract and compounded at the risk-free rate until the settlement date.

Prior researchers have confronted the challenge of developing a forecasting model which can be integrated into the options pricing framework while, at the same time, providing accurate estimates and forecasts. These models can be categorised into two distinct groups. The first group relies on the standard Stochastic Brownian Motion framework similar to the approach proposed by Hull and White (1990). Studies from this group include those of Dischel (1998a, 1998b, 1999) that were later improved on by Alaton, Djhiche and Stillberger (2002). The second group which encompasses the work of Cao and Wei (2003) and Campbell and Diebold (2000), focus on the use of time series analysis.[6]

The study by Dischel (1998b) pioneered the use of Stochastic Brownian Motion as an innovation generating mechanism within a temperature forecasting model. He employed an approach that was similar to Hull and White's (1990) pricing model for interest rate derivatives. The pricing model proposed by Dischel (1998b) is established under a simple, two parameter stochastic differential equation:[7]

$$dT = \alpha[\theta(t) - T(t)]dt + \gamma\tau(t)\, dm_1 + \delta\sigma(t)dm_2 \qquad (6)$$

T is defined as temperature and t as time. θ is a measure of the average daily historical measure of temperature over time, while τ and σ represent the distribution of T and ΔT, respectively.[8] ΔT signifies the one period change of T. Using the technique of finite differencing, equation (6) is reduced to a one-parameter discrete analogue that is given by equation (7) below.

$$\hat{T}_{n+1} = \alpha\theta_{n+1} + \beta\hat{T}_n + \delta\Delta T_{n,n+1} \qquad (7)$$

$\hat{T}_{n+1}$ is an estimate of the projected temperature for period n+1, while $\Delta T_{n,n+1}$ represents the variation in temperature between period n and n+1. The parameter δ is restricted to be less than or equal to one, as is the sum of α and β. The parameter, θ_n, is the value to which the simulated temperature reverts to in the absence of randomness. The value of θ_n is calculated for each date through the averaging of all temperatures on the same day from historical data using the following equation (8):

$$\theta_n = \frac{\sum_{i=1}^{N} T_{i,n}}{N}, \qquad (8)$$

where N is the number of years of observations, and n the specific date which takes a value between 1 and 365.

Dornier and Queruel (2000) argued that direct adaptation of the Hull and White model, as utilized by Dischel (1998a, 1998b), is undesirable because the mean value of simulated temperature from the Hull and White model deviates from the historical mean. Further, they demonstrated the need to include the term $d\theta_t$ in the adapted model given by equation (9) below:

$$dT_t = d\theta_t + a_t\,(\theta_t - T_t)\,dt + \sigma_t\,dW_t \qquad (9)$$

Later studies improved on Dischel's (1998b) model by incorporating Dornier and Queruel's (2000) suggestion, as well as mean-reversion and seasonality effects. In the study by Alaton, Djhiche and Stillberger (2002), the seasonality effect was modelled by a linear oscillation model which follows a sine function. However, the use of sinusoids to control for the seasonality effect was discouraged by Moreno (2000) who argued that the regularity of the sinusoids prevented an adequate fit of the asymmetric evolution of temperature in summer relative to winter. Furthermore, Moreno (2000) demonstrated that the distribution of the noise is not homogenous throughout time. This induces a bias in the out-of-sample forecasts despite a satisfactory goodness of fit.[9] This aspect is controlled for in the Alaton, Djhiche and Stillberger's (2002) model by relaxing the constant variance assumption via the use of month-specific variances.

As an alternative method of temperature forecasting, Cao and Wei (2003) proposed the use of a time series approach which results in an ARMA model with periodic variance. Their model incorporated the sine wave function to model the seasonality effect. The criticism put forward by Moreno (2000) regarding the regularity of sinusoids, still persists with their time-series forecasting model. Additionally, the historical average is adjusted for long-term trend to accommodate global warming effects as well as abnormally cold or warm years. Despite its ability to accommodate most of the required features of temperatures, Cao and Wei (2003) argued against the use of a diffusion process, especially a mean-reverting process. They pointed out that a one-factor diffusion process is unable to incorporate autocorrelations in temperature innovations beyond a lag of one period.

Recently, Campbell and Diebold (2001) proposed an extension to the conventional ARMA model utilized by Cao and Wei (2003). Their model also incorporated both long-term trend and seasonality. Seasonality is controlled for by utilising Fourier transforms. Additionally, the autoregressive lag structure is based on both the Akaike Information Criteria (AIC) and Schwarz Information Criteria (SIC). Their documented results, based on both these approaches, were consistent.

III. Data

The data set utilized in this study covers the period from 1st January, 1979, to 31th December, 2002, covering contracts for all ten cities listed on the Chicago Mercantile Exchange (CME). The data set consists of the daily minimum, maximum and average temperature, as well as the daily and monthly-accumulated HDD and CDD indices.[10]

Table 2 compares the descriptive statistics of the daily average temperature, monthly-accumulated HDD and monthly-accumulated CDD for all ten cities that comprised the sample. Consistent with Dischel (1998b) and Cao and Wei (2003), the evidence presented in Table 2 indicates disparities in the average temperature experienced by different cities in the United States.

<INSERT TABLE 2 ABOUT HERE>

Figure 1 graphs the average daily temperature for the three year period between 1st January, 1984, and 31st December, 1986, for Atlanta and Chicago. Consistent with prior studies, the daily average temperature is seasonal and appears to follow a sine-wave pattern. For clarity of presentation, the graphical analysis in Figure 1 focuses on Atlanta (ATL), while analysis on Chicago (ORD) is presented solely for the purpose of comparison. After observing this randomly selected set of 3-year periods for each city, it is apparent that the daily average temperature exhibits a full seasonal cycle of one calendar year. This seasonal behaviour is consistent throughout the full observation period from 1979 to 2002.

<INSERT FIGURE 1 ABOUT HERE>

IV. Methodology

Six different approaches to temperature forecasting are examined in this article. The first three approaches utilize Monte Carlo simulations, while the last three incorporate models established under an autoregressive moving average (ARMA) process. For each of the models that rely on an ARMA process, 5000 forecast realisations were simulated using an error distribution extracted from the historical data.[11] The forecasts reported for these models are represented by the expected values of the 5000 forecast realisations generated for each time period over the forecast horizon. The six models are defined below.

- $Model_1$ represents a naïve temperature forecasting model which relies on the distribution of the historical average temperature. Simulations of 5000 iterations each are generated for each city and for each month. A list of the resultant distributions is reported in Panel A of Table 3.
- $Model_2$ follows the methodology proposed by Dischel (1998b) as outlined by equation (6) to (8) in Section II. Parameter estimates are reported in Panel B of Table 3.
- $Model_3$ reports the forecasted temperature based on Alaton et. al (2002) and outlined by equation (A.1) to (A.5) in the Appendix. The parameter estimates are reported in Panel C of Table 3.

<INSERT TABLE 3 ABOUT HERE>

- *Model*$_4$ reports the forecasted temperature based on a conventional ARMA model without controlling for long-term trend and seasonality. Following an ARMA(p,q) process, the temperature at period t is given as follows:

$$T_t = \sum_{p=1}^{k} \rho_p T_{t-p} + \sum_{q=1}^{l} \sigma_q * \varepsilon_{t-q} \qquad (10)$$

  The parameter estimates are reported in Panel D of Table 3.

- *Model*$_5$ follows the approach of Cao and Wei (2003). Derivations of the model are reported in equation (A.6) to (A.10) in the Appendix. The parameter estimates are reported in Panel E of Table 3.

- *Model*$_6$ relies on Campbell and Diebold's (2001) methodology as described by equations (A.11) to (A.14) in the Appendix.[12] Parameter estimates for this model are found in Table 4.

In order to compare the models outlined above, both in-sample and out-of-sample tests are performed. Recall that the in-sample period is the estimation period, while the out-of-sample period covers the forecasting horizon. A five-year period covering the period between 1st January, 1998, and 31st December, 2002, is reserved for out-of-sample testing as this period covers the time from commencement of trading on the CME. Temperature data between the period of 1979 and 1998 is utilized to estimate the parameters of the models that generate the forecasts for the out-of-sample period.[13] The parameter estimates reported in Tables 3 are based on analysis within the in-sample period.

Weather derivatives are predominantly written in terms of the monthly-accumulated sum of the daily HDDs and CDDs and can be written and traded months prior to the settlement date. Therefore, market-makers and other participants are required to forecast in advance estimates of both the AccHDD and AccCDD indices. To address this issue, the forecast accuracy of the forecasted indices was tracked over the out-of-sample period. Unless specified, the forecasted indices for the out-of-sample period were generated using a dynamic forecasting approach for up to one year or 365 days-ahead. For this approach, the forecast for the t+1 period was generated using the in-sample information included at period t. The forecasts for t+2 and subsequent time periods over the forecasting horizon, however, are generated conditional on all information up to and including period t, as well as forecasted values extending from period t+1.

To examine forecast accuracy, forecasting error is computed as:

$$FE_{t,i} = Actual_t - Forecast_{t,i} \ , \tag{11}$$

where $FE_{t,i}$ is the forecast error from model *i* at time *t*, and is given by the deviations between the actual and the forecasted values. Obviously, examination of the forecast accuracy is not limited to the daily average temperature, but is also extended to the monthly-accumulated HDD and CDD. Analysis on the monthly-accumulated HDD covers the months of January, February, March, April, October, November and December, while the analysis on the monthly-accumulated CDD covers the months of April, May, June, July, August, September and October. As outlined in Panel B of Table 1, these months are the months for which these contracts are traded.

Three measures of forecasting error are utilized in this study, namely, RE (Relative Error), Theil U-statistics and UAPE (Unbiased Absolute Percentage Error).

$$RE = \frac{FE_{t,i}}{Actual_t} \tag{12}$$

$$Theil = \frac{RMSE_i}{RMSE_1} \tag{13}$$

$$UAPE = \left| \frac{Actual_t - Forecast_{t,i}}{\left(\frac{Actual_t + Forecast_{t,i}}{2} \right)} \right| \tag{14}$$

The use of Relative Error (RE) is limited to the examination of a model's propensity to over- and under-forecast the actual value. A positive RE indicates that a forecasting model under-forecasts, while a negative RE indicates a model over-forecasts. Theil's U-statistic is utilized to examine the forecasting ability of each model relative to the naïve model, $Model_1$. Theil's U-Statistic is the ratio of the Root Mean Squared Error (RMSE) for each model to that of $Model_1$. The last measure of forecasting error, UAPE, circumvents criticism of the conventional Mean Squared Error (MSE) method of assessing forecast error and model comparison on the basis of the scale of the error [see Chatfield (1988), Armstrong and Collopy (1992), Fildes (1992)]. Makridakis (1993) proposed the use of the UAPE measure as an unbiased measurement with respect to the scale of the error.

V. Results

Recall that $Model_1$ is a naïve temperature forecasting model while $Model_2$ is based on the methodology proposed by Dischel (1998b). $Model_3$ also relies on Monte Carlo simulations but reports forecasted temperature based on the approach of Alaton et al. (2002). $Model_4$, $Model_5$ and $Model_6$ are ARMA-type time series models that all differ with regard to their treatment of seasonality.

<INSERT TABLE 4 ABOUT HERE>

Table 4 reports the descriptive statistics of actual and forecasted values for both in-sample and out-of-sample periods of the daily average temperature and the monthly-accumulated HDD (AccHDD) and CDD (AccCDD) indices for all models. For each model, we report the summary statistics aggregated over all ten cities for which weather derivatives are listed on the CME.

<INSERT TABLE 5 ABOUT HERE>

After examination of the means of the in-sample Relative Error (RE), averaged over all ten cities and reported in Table 5, we concluded that all models (with the exception of $Model_6$) significantly under-forecast the AccHDD and AccCDD indices, as well as the daily average temperature. $Model_6$ was found to over-forecast the AccHDD Index, but not the AccCDD Index and the daily average temperature.

The out-of-sample tests provide mixed results. All models under-forecast the daily average temperature. With the exception of $Model_4$ and $Model_5$ that significantly under-forecast, all other models over-forecast the AccHDD Index. The converse is the case for the AccCDD Index. $Model_4$ and $Model_5$ were found to under-forecast that index, while the remainder over-forecast. Again, the mean RE corresponding to $Model_4$ and $Model_5$ is statistically significant, which is in contrast to that of the other models.

<INSERT FIGURE 2 ABOUT HERE>

<INSERT FIGURE 3 ABOUT HERE>

Figures 2 and 3 graphically depict the mean UAPE of all six different forecasting models for the in-sample and out-of-sample periods, respectively. In both the in-sample and out-of-sample periods, significantly higher mean UAPEs are associated with the AccCDD Index. This suggests greater difficulty in forecasting the AccCDD Index relative to both the AccHDD Index and the daily average temperature. This finding is consistent with Roustant, Laurent, Bay and Carraro (2001) who find a greater price uncertainty for the CDD months. Those authors attribute their finding to the same choice of threshold level (65^{o}F) in the calculation of both the HDD and CDD indices, and the different accumulation effect it has during the winter and summer months.

For the in-sample period specifically, Figure 2 demonstrates that models that rely on an ARMA process had lower UAPE's for the AccHDD Index than the daily average temperature, while the UAPE's for models that rely on Monte Carlo simulation had higher values. Additionally, Figure 2 also demonstrates that models that rely on an ARMA process had lower UAPE's for both indices and the daily average temperature than those based on Monte Carlo simulation.

For the out-of-sample period as depicted in Figure 3, all models were characterized by a higher UAPE for the forecasted AccHDD and AccCDD indices than for the forecasted daily average temperature. However, the relatively better goodness-of-fit of the ARMA-type models when compared to the others in the in-sample period, was reduced markedly out-of-sample. This is consistent with Moreno (2000), who states that the heterogeneity of the noise distribution induces a bias in the out-of-sample forecasts while providing an adequate goodness-of-fit for the in-sample period.

After comparing the in-sample UAPE's of the AccHDD and AccCDD indices for $Model_1$ with those from $Model_2$ in Figure 2, it was clear that the forecasting error for each index generated from $Model_2$ is higher than that of $Model_1$. Additionally, relative to $Model_1$, the estimation error from fitting $Model_2$ to the daily average temperature was only marginally improved. A consistent finding to that above was observed from Figure 3 for the out-of-sample period. We concluded that, after controlling for long-term trend as proposed by Dischel (1998b), forecast accuracy was not significantly improved. Controlling for seasonality however, generated some improvement in the quality of the forecast for the AccCDD Index. This was evident when comparing the UAPE's for the AccCDD Index from $Model_3$ with that of $Model_2$ for models that rely on Monte Carlo simulation, and by a similar comparison of the UAPE's from $Model_5$ and $Model_6$ with that of $Model_4$ for models that are based on an ARMA process.

From Figure 2, we observed that the direct ARMA model ($Model_4$) provided better goodness-of-fit across the in-sample period for all three variables. However, from Figure 3, $Model_5$, appears to be the best model for forecasting the AccHDD Index out-of-sample. $Model_5$ is the model proposed by Cao and Wei (2003) which is based on an ARMA process and incorporates both long-term trend as well as correcting for seasonality. When forecasting the AccCDD, the naïve model, $Model_1$, was found to have the lowest UAPE. For the daily average temperature, $Model_3$, $Model_4$, $Model_5$ and $Model_6$ are found to exhibit an equal level of forecasting error.

<INSERT FIGURE 4 ABOUT HERE>

<INSERT FIGURE 5 ABOUT HERE>

Figures 4 and 5 refer to the monthly in-sample UAPE measures for the monthly-accumulated HDD (AccHDD) Index and the monthly-accumulated CDD (AccCDD) Index, respectively. Consistent with Roustant, Laurent, Bay and Carraro (2001), the monthly UAPE's reported in Figure 4 and 5 were higher for the transition months, namely April and October. This finding is to be anticipated. Due to the method of calculating both the daily HDD and CDD by truncating the temperature at 65^{o} F, the error distribution of the daily HDD and CDD will no longer be white noise. Hence, the greater the likelihood of the temperature crossing the threshold level, the greater will be the error generated by the forecast. This explains the higher UAPE documented for the transition months, where the average daily temperature is most likely to be close to the threshold level. Models that rely on the ARMA process exhibited a better in-sample fit relative to those that rely on the Monte Carlo simulation.

<INSERT FIGURE 6 ABOUT HERE>

<INSERT FIGURE 7 ABOUT HERE>

Figure 6 and 7 illustrate the UAPE of all six models partitioned by month for the out-of-sample period. While the results reported in Figure 6 and 7 still exhibited relatively higher forecasting error during the transition months, the evidence was less pronounced when compared to the in-sample period. Additionally, while models that relied on an ARMA process consistently exhibited an improved goodness-of-fit for the in-sample period relative to those that relied on Monte Carlo simulation, the results from the out-of-sample period were mixed. The results illustrated in Figure 6 and 7 demonstrated that the best forecasting model varied across months.

<INSERT TABLE 6 ABOUT HERE>

Table 6 reports the best forecasting model for each U.S. city. For each city, the best model was defined to be the one that generated the lowest UAPE. Importantly, the preferred forecasting model varied between cities. Further, the best model (based on the UAPE) for the in-sample months is not consistent with that for the out-of-sample months. This could be driven by the disparity in temperature behaviour between cities, as previously documented by Cao and Wei (2003) and Campbell and Diebold (2001) and reported in Table 2. Alternatively, it could simply be the absence of a general model which adequately explains the temperature behaviour of all ten cities. The monthly UAPE measures, as illustrated in Figures 4 to 7, indicated that the model which best forecasted the monthly-accumulated HDD and CDD varied between months within the estimation period, and over the forecast horizon.

<INSERT FIGURE 8 ABOUT HERE>

<INSERT FIGURE 9 ABOUT HERE>

<INSERT FIGURE 10 ABOUT HERE>

Given the nature of weather derivative contracts that represent the monthly accumulated sum of the daily HDD or CDD index, market participants are required to forecast these indices months in advance. Figures 8 to 10 compare the forecast errors for the two indices and the daily average temperature that were generated by $Model_3$ to $Model_6$ and based on dynamic one (*M1*) and three (*M3*) months-ahead forecasts. Given the design of $Model_1$ and $Model_2$, dynamic step-ahead forecasts were not considered. The results demonstrated that the one month-ahead forecasts for all models provided improved forecasting ability relative to those forecasts over the longer time horizon. As illustrated in Figures 9 and 10, the one month-ahead forecasts for the AccCDD Index and the daily average temperature resulted in better quality forecasts than those three months-ahead. These results were less pronounced for the AccHDD Index depicted in Figure 8.

Up to this point, comparative analyses have been performed relative to the naïve model, $Model_1$, using the settlement level as the benchmark. In order to fairly evaluate and compare the forecasting models, it is important to compare the estimated price generated by each model against the market price of the contract. Based on this analysis, the results would provide an insight into the ability of models to forecast the settlement level relative to that of the market. In order to perform the analysis, 212 CDD futures contracts and 158 HDD futures contracts traded on CME between 1st January, 1998, and 31st December, 2002, were extracted from *Reuters*.[14] For each contract, we obtained information on the type of the contract, the date the contract was traded, the settlement date of the contract and the price of the contract. Temperature data up to and including the date the contract was traded was used to forecast the price of the contract. The RMSE of each of the forecasts relative to the settlement level was computed. Similarly, the RMSE of the traded price relative to the settlement level was also computed. The comparison was performed using the *Theil U-Statistic* illustrated in equation (13), with the RMSE of the traded price utilized as the denominator. This approach compared the accuracy of the forecasting models relative to the traded price. A *Theil* value below 1 signified that the price generated using the forecasting models was more accurate relative to those traded on the market. Conversely, a *Theil* value above 1 inferred that the price generated by the models was less accurate than those traded on the market.

<INSERT FIGURE 11 ABOUT HERE>

<INSERT FIGURE 12 ABOUT HERE>

The out-of-sample results illustrated in Figure 11 demonstrated that all six models were able to provide a more accurate forecast of the true value of the AccHDD contract relative to the market. With the exception of perhaps the November contracts, the forecasted settlement levels generated by all six models were more accurate than those inferred by the market. However, the results from the AccCDD contracts given in Figure 12 show that market price was a more accurate representation of the settlement level.

<INSERT FIGURE 13 ABOUT HERE>

<INSERT FIGURE 14 ABOUT HERE>

Similar results are documented in Figures 13 and 14 where the analysis was performed on city-by-city basis. From Figure 13, with the exception of contracts that are traded for Dallas and New York, all six forecasting models examined in this study were able to provide significantly better forecasts of the AccHDD Index. On the other hand, from Figure 14 we observed that for the AccCDD Index, the forecasts generated by the models were generally less accurate when compared to the traded price. This finding confirmed that the models examined in this study, while able to adequately forecast the AccHDD Index, generated a forecast of relatively poor quality when utilized to estimate the AccCDD Index. This could have resulted from the use of the risk-free rate as a discount rate when, in fact, a non-zero risk premium existed in this market.[15]

VI. Conclusions

After examining six different temperature forecasting models, no one model was able to constantly outperform the others. Relative to models that relied on Monte Carlo simulations, relying on an ARMA process produced a better goodness-of-fit for the in-sample estimates. However, the superiority of these forecasts was not reflected in the out-of-sample period. For the in-sample period, all models with the exception of $Model_6$, the model proposed by Campbell and Diebold (2001), exhibited a tendency to under-forecast all three underlying variables. For the out-of-sample period, models that relied on Monte Carlo simulations were more likely to over-forecast the AccHDD Index and under-forecast the AccCDD Index. For the models that relied on an ARMA process, the conventional model ($Model_4$) and the Cao and Wei (2003) model ($Model_5$) both under-forecasted the AccHDD. Conversely, *$Model_6$* over-forecasted the AccHDD. The exact opposite result held for the AccCDD Index. Both types of models, however, exhibited a tendency to under-forecast the daily average temperature.

The accuracy of forecasts, certainly beyond 30 days (one month), is unreliable. Controlling for seasonality tends to increase forecast quality but controlling for long-term trend did not significantly improve forecast accuracy. It also appeared more difficult to forecast the AccCDD Index than the AccHDD Index. The most appropriate forecasting model varied between cities and months as well as across the estimation and forecasting samples. The forecasting models examined in this study were able to provide a more accurate forecast of the price of the AccHDD futures contracts than the actual traded price. However, relative to the market, these models performed poorly when utilized to price the AccCDD futures contracts.

This study is limited to examining the temperature forecasting models that have been proposed by prior studies as the underlying models for valuing weather derivatives contracts. It is important to reiterate that the temperature forecasting models examined in this study are not directly comparable to models utilized by meteorological bureaus around the world for their temperature forecasts. Therefore, the results and conclusions derived in this study need to be interpreted with cautious.

Appendix.

This section details the steps undertaken in order to generate the forecasts analysed in this study. For the naïve model, $Model_1$, temperature is estimated by simulating from a temperature distribution obtained from historical data. That is, the distribution of temperature for each month is estimated using the temperature data up to 31st December 1997. This distribution is then utilized to simulate the temperature for the year 1998. Similarly, temperature data up to 31st December 1998 is used to simulate the temperature for the year 1999 and so on.

The second model, $Model_2$, replicates the method of Dischel (1998b). Following Dischel (1998b), equation (7) is estimated and the best estimates are obtained using trial and error. For each city, we run equation (7) using different combinations of α, β and δ, where all three parameters take the value between 0 and 1. Consistent with Dischel (1998b), we restricted α and β, so that the sum of the two parameters is equal to 1. The model is estimated by taking the value of $\alpha = 0.1$, $\beta = 0.9$ and $\delta = 0.1$. The model is then re-estimated by taking $\alpha = 0.2$, $\beta = 0.8$ and $\delta = 0.1$. All combinations of α, β and δ are estimated and the one with the lowest forecasting error is used as the specified model.

The third model, $Model_3$, is introduced by Alaton et. al (2002), which extends Dischel's (1998b) temperature-forecasting model. Alaton et. al (2002) proposed the use of the Ornstein-Uhlenbeck process to model the average temperature in order to obtain a process that reverts to the mean as proposed by Dornier and Queruel (2000). In order to estimate $Model_3$, the following regression was run:

$$Y_t = a_1 + a_2 t + a_3 \sin(\varpi t) + a_4 \cos(\varpi t) \qquad \text{(A.1)}$$

where Y_t denotes the daily average temperature of the in-sample period, while t takes a value between 1 and 365. ϖ is given by $2\pi/365$. The mean temperature at time t is estimated as follows:

$$T_t^m = A + Bt + C\sin[(2\pi/365)t + \varphi] \qquad \textit{(A.2)}$$

where A=a_1, *B* = a_2, *C* = $\sqrt{a_3^2 + a_4^2}$ and $\varphi = \arctan\left(\frac{a_4}{a_3}\right) - \pi$.

Then, the following process is estimated:

$$dT_t = \left\{\frac{dT_t^m}{dt} + a(T_t^m - T_t)\right\}dt + \sigma_t dW_t \qquad (A.3)$$

where

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu - 2}\sum_{j=0}^{N_\mu - 1}\left(\tilde{T}_j - \hat{a}T_{j-1}^m - (1-\hat{a})T_{j-1}\right)^2 \qquad (A.4)$$

$$\hat{a} = -\log\left(\frac{\sum_{i=1}^{n}\frac{T_{i-1}^m - T_{i-1}}{\sigma_{i-1}^2}\left\{T_i - T_i^m\right\}}{\sum_{i=1}^{n}\frac{T_{i-1}^m - T_{i-1}}{\sigma_{i-1}^2}\left\{T_{i-1} - T_{i-1}^m\right\}}\right) \qquad (A.5)$$

φ is included in order to control for annual minimum and maximum mean temperatures that do not occur at January 1 and July 1 respectively. *a* represents the speed of mean reversion, while μ, which represents a specific month, is included in order to control for month specific volatility. Applying the temperature forecasting model above on the option-pricing model leads Alaton et. al (2002) to conclude that the proposed model exhibits a greater degree of desirability relative to a straight Monte Carlo simulation.

Model$_4$ is estimated by using a direct ARMA (Autoregressive Moving Average) approach. Following Cao and Wei (2003), an ARMA (3,3) model is specified. *Model*$_5$ represents an alternative approach to temperature forecasting. It relies on the use of an ARMA (Autoregressive Moving Average) model as proposed by Cao and Wei (2003). Their rationale for this approach is based on the inability of a one-factor diffusion process to incorporate autocorrelation into temperature innovations for lags beyond one. The de-meaned and de-trended residual of the daily temperature, U_t, is given by the difference between the actual temperature, $Y_{,t}$, and the de-meaned and de-trended historical daily average temperature, $\overline{Y}_t$:

$$U_t = Y_t - \left(\frac{\beta}{365} \left(t - T_t \right) \right) + \overline{Y}_t \qquad \text{t = 1,2,…,T} \qquad (A.6)$$

$$U_t = \sum_{i=1}^{k} \rho_i U_t + \sigma_t * \xi_t \qquad yr = 1,2,…,22 \qquad (A.7)$$

$$\sigma_t = \sigma_0 - \sigma_1 \,|\sin(\pi t / 365 + \varphi)| \qquad (A.8)$$

$$\xi_t \sim i.i.d \ \mathrm{N}(0,1) \qquad (A.9)$$

$$\forall \quad yr = 1,2,…,22 \text{ and } t = 1,2,…,365. \qquad (A.10)$$

Model$_6$ represents a similar approach that was suggested by Campbell and Diebold (2001). A Fourier transform is used to control for seasonality in temperature. The choice of number of lags to be included into the model is selected based on the AIC (Akaike Information Criteria) and SIC (Schwarz Information Criteria).

$$T_t = Trend_t + Seasonal_t + \sum_{l=1}^{L} \rho_{t-1} T_{t-1} + \sigma \varepsilon_t \qquad (A.11)$$

$$Trend_t = \beta_0 + \beta_1 t \qquad (A.12)$$

$$\mathrm{Seasonal_t} = \sum_{p=1}^{P} \left(\delta_{c,p} \cos\left(2\pi p \frac{d(t)}{365} \right) + \delta_{s,p} \sin\left(2\pi p \frac{d(t)}{365} \right) \right) \qquad (A.13)$$

$$\varepsilon \overset{\text{iid}}{\sim} (0,1) \qquad (A.14)$$

The forecasts generated based on the models that rely on an ARMA process, namely, *Model*$_4$, *Model*$_5$ and *Model*$_6$, represent the expected value of the 5,000 forecast paths created. For each model, the error distribution of the in-sample period is estimated. We then create 5,000 forecast paths using errors that are simulated based on the obtained distribution.

References

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

Björk, T., 1998, "Arbitrage Theory in Continuous Time", Oxford University Press.

Brennan, M.J. and E. Schwartz, 1982, "An equilibrium model of bond pricing and a test of market efficiency", *Journal of Financial and Quantitative Analysis* 17, 75 – 100.

Brody, D.C., 2000, "Modern Mathematical Theory of Finance", Nippon-Hyoronsya, Tokyo.

Brody, D.C., J. Syroka, and M. Zervos, 2001, "Dynamical Pricing of Weather Derivatives", *Quantitative Finance* 2, 189 – 198.

Campbell, S. and F.X. Diebold, 2001, "Weather Forecasting for Weather Derivatives", *PIER Working Paper 01 – 031*, Penn Institute for Economic Research, University of Pennsylvania.

Cao, M. and J. Wei (2003), "Weather Derivatives Valuation and Market Price of Weather Risk", Manuscript, York University and University of Toronto.

Challis, S., 1999, "Bright Forecast for Profits", *Reactions*, June edition.

Chan, K., A. Karoly, F. Longstaff, and A. Sanders, 1992, "An empirical comparison of alternative models of the short term interest rate", *Journal of Finance* 3, 1209 – 1226.

Chatfield, C., 1988, “Apples, oranges and mean square error”, *International Journal of Forecasting* 4, 515 – 518.

Chicago Mercantile Exchange Rulebook.

Collopy, F. and J.S. Armstrong, “Another Error Measure for Selection of the Best Forecasting Method: The Unbiased Absolute Percentage Error”, http://hops.wharton.upenn.edu/forecast/paperpdf/armstrong0unbiasedAPE.pdf.

Cox, J.C., J. Ingersoll, and S. Ross, 1985, “A theory of the term structure of interest rates”, *Econometrica* 53, 385 – 407.

Davis, M., 2001, “Pricing Weather Derivatives by Marginal Value”, *Quantitative Finance* 1, 305 – 308.

Dichel, B., 1998a, “Black-Scholes won’t do”, *Risk*, October Edition, 8 – 9.

Dichel, B., 1998b, “At last: a model for weather risk”, *Risk, Energy and Power Risk Management*, March edition, 20 – 21.

Dichel, B., 1999, “Shaping History for Weather Risk Management”, *Special “Weather Report” of Risk, Energy and Power Risk Management*, August edition.

Dornier, F., and M. Queruel, 2000, “Caution to the Wind”, *Risk, Energy and Power Risk Management*, August edition.

Geman, H., (1998), *Insurance and Weather Derivatives: From Exotic Options to Exotic Underlyings,* London: Risk Publications.

Hanley, M., 1999, “Hedging the Force of Nature”, *Risk Professional* 1, 21 – 25.

Hull, J.C. and A. White, 1990, “Pricing interest rate derivatives”, *Review of Financial Studies* 3 (4), 573 – 592.

Hunter, R., 1999, “Managing Mother Nature”, *Derivatives Strategy*, February edition.

Makridakis, S., 1993, “Accuracy measures: Theoretical and practical concerns”, *International Journal of Forecasting* 9, 527 – 529.

Moreno, 2000, “Riding the Temp”, *Weather Derivatives, FOW Special Supplement*, December.

Nicholls, M., 1999, “Winter warms weather treaders”, *Environmental Finance*, October edition.

Roustant, O., J.P. Laurent, X. Bay, L. Carraro, 2001, “An Empirical Study of The Price Uncertainty of Some Weather Derivatives”, Working Paper, École des Mines de Saint-Etienne and ISFA, Université Lyon I.

Torró, H., V. Meneu, E. Valor, 2001, “Single Factor Stochastic Models with Seasonality Applied to Underlying Weather Derivatives Variables”, Working Paper, Universitat de València.

Vasicek, O., 1977, “An equilibrium characterization of the term structure”, *Journal of Financial Economics* 5, 177 – 188.

We thank Graham Partington, Alex Frino, Maurice Peat, Adam Kucera, Craig McBurnie, Jason West, Joel Fabre, an anonymous referee, the Editor and the participants of the Capital Markets CRC Seminar for their comments. We also thank Ian Young and Ian Sue for programming assistance. The authors acknowledge the Securities Industry Research Centre of Asia-Pacific (SIRCA) for providing access to the Reuters and Data Stream. This research is funded by the CRC for Technology Enabled Capital Markets.

[1] Torró, Meneu and Valor (2001) report derivatives written on such non-tradable assets as short term interest rates and weather derivatives.

[2] For further details see Brody, Syroka, Zervos (2001).

[3] Indeed, the market for weather derivatives has been categorized as one of the classic examples of an incomplete market (see Björk ,1998, and Brody ,2000).

[4] Higher volatility and bid-ask spread represents higher cost of trading which, in turn, reduces the supply of the liquidity (Dornier and Queruel, 2000).

[5] 10401.A., Chapter 104, CME Rulebook.

[6] See the Appendix to this paper for a more detailed exposition of these models.

[7]A similar process is also employed by Davis (2001) and Geman (1999).

[8] Originally, Dischel (1998a) sample the two distributions using Wiener processes, but this approach was later replaced in Dischel (1998b) by sampling from an actual distribution.

[9] A Similar conclusion is inferred by Roustant, Laurent, Bay and Carraro (2001).

[10] The data set was obtained from the CME.

[11] Details of the probability distributions fitted to the errors from the models are available from the authors on request.

[12] Campbell and Diebold (2001) utilize two different model selection methods, AIC (Akaike Information Criteria) and SIC (Schwartz Information Criteria). Given that the authors infer that both selection criteria produce consistent results, only the forecasts from models selected on the basis of AIC are analysed in this article.

[13] To ensure consistency, all observations on 29th February are deleted from the sample.

[14] The Reuters data is provided by SIRCA the Securities Industry Research Centre of Asia-Pacific.

[15] Ideally, we would have preferred to perform our analysis using the actual discount rate employed in the market. However, the unavailability of this data prevented such an option.

Table 1. CME Contract Symbols, Measuring Stations and Contract Specifications

Panel A: CME Contract symbols and measuring stations

City	CDD Symbol	HDD Symbol	Measuring Station (Symbol)	Automated Weather Station
Atlanta	K1	H1	ATL	Atlanta Hartsfield Airport weather station
Chicago	K2	H2	ORD	Chicago O'Hare Airport weather station
Cincinnati	K3	H3	CVG	Cincinnati – Northern Kentucky International Airport
Dallas	K5	H5	DFW	Dallas Forth Worth Airport
Des Moines	K9	H9	DSM	Des Moines International Airport
Las Vegas	K10	H10	LAS	Las Vegas McCarran International Airport
New York	K4	H4	LGA	New York La Guardia Airport
Philadelphia	K6	H6	PHL	Philadelphia International Airport
Portland	K7	H7	PDX	Portland International Airport
Tucson	K8	H8	TUS	Tucson International Airport

Panel B: Weather derivatives contract specification

	Futures contract
Contract size:	\$100 times the respective CME HDD/CDD index
Minimum tick size:	1.00 HDD/CDD index point = \$100
Contract Month:	Oct, Nov, Dec, Jan, Feb, Mar and Apr for HDD. Apr, May, Jun, Jul, Aug, Sep and Oct for CDD.
Termination day:	9:00 am (Chicago time) on the first exchange business day that is at least two calendar days after the contract month
Final settlement day:	The first exchange business day that is at least two calendar days after the contract month
Final settlement price:	The exchange will settle the contract to the respective CME HDD/CDD Index of the contract month as calculated by EarthSat
Trading hours:	24 hour access to Globex electronic-trading system

Table 2 Comparison of Descriptive Statistics of Daily Average Temperature, Monthly-Accumulated HDD and Monthly-Accumulated CDD for all Ten U.S. Cities

	ATL	ORD	CVG	DFW	DSM	LGA	PHL	PDX	TUS	LAS
Avg T Mean	62.67	53.87	65.84	50.44	67.98	55.53	49.64	54.38	55.61	69.61
Avg T Median	64.00	55.50	67.00	52.00	67.25	56.00	50.50	53.50	56.00	69.50
Avg T StdDev	15.02	18.36	16.33	21.25	16.90	17.20	21.12	11.55	17.50	15.88
AccHDD Mean	387.89	702.27	338.33	867.88	309.44	640.60	864.18	552.19	645.84	213.33
AccHDD Median	401.75	703.75	324.25	873.25	303.25	665.25	868.50	551.00	660.25	204.50
AccHDD StdDev	219.88	297.55	218.80	369.22	205.53	271.49	333.14	176.40	274.94	144.78
AccCDD Mean	268.46	152.91	377.01	148.31	463.57	171.63	121.98	57.80	178.90	447.05
AccCDD Median	259.00	120.50	389.50	114.00	496.00	149.00	81.25	34.25	142.50	495.50
AccCDD StdDev	171.99	129.61	214.17	137.26	264.87	149.84	124.54	65.20	153.78	211.98

Legend:
Avg T = Daily Average Temperature
AccHDD = Monthly-accumulated HDD Index
AccCDD = Monthly-accumulated CDD Index

Table 3. Parameter Estimates

	ATL	CVG	DFW	DSM	LAS	LGA	ORD	PDX	PHL	TUS
Panel A: Model$_1$										
Jan	Weibull	Weibull	Weibull	Weibull	Gamma	Weibull	Weibull	Weibull	Weibull	Weibull
Feb	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull
Mar	Weibull	Erlang	Weibull	Erlang	Gamma	Erlang	Erlang	Weibull	Erlang	Weibull
Apr	Weibull	Chi-Sq	Weibull	Weibull	Gamma	Gamma	Erlang	Gamma	Erlang	Weibull
May	Weibull	Weibull	Gamma	Erlang	Weibull	Gamma	Gamma	Gamma	Erlang	Weibull
Jun	Weibull	Weibull	Erlang	Gamma	Weibull	Gamma	Weibull	Erlang	Weibull	Weibull
Jul	Erlang	Gamma	Weibull	Gamma	Weibull	Gamma	Erlang	Erlang	Gamma	Erlang
Aug	Erlang	Weibull	Weibull	Weibull	Weibull	Erlang	Gamma	Gamma	Weibull	Erlang
Sep	Weibull	Weibull	Weibull	Gamma	Weibull	Gamma	Gamma	Gamma	Beta	Weibull
Oct	Weibull	Gamma	Weibull	Weibull	Weibull	Erlang	Gamma	Gamma	Erlang	Weibull
Nov	Weibull	Erlang	Gamma	Weibull	Gamma	Erlang	Gamma	Weibull	Gamma	Weibull
Dec	Weibull	Weibull	Weibull	Weibull	Gamma	Weibull	Gamma	Weibull	Weibull	Weibull
Panel B: Model$_2$										
α	0.30	0.10	0.40	0.10	0.10	0.20	0.30	0.30	0.30	0.40
β	0.70	0.90	0.60	0.90	0.90	0.80	0.70	0.70	0.70	0.60
δ	0.60	0.40	0.50	0.30	0.10	0.40	0.40	0.70	0.50	0.10
Panel C: Model$_3$										
A	62.89	52.79	66.72	49.64	71.61	54.46	48.17	57.30	54.60	72.40
B	0.00	0.00	-0.01	0.00	-0.02	0.00	0.01	-0.02	0.00	-0.02
C	18.80	22.81	20.57	27.15	22.91	22.17	25.10	15.13	22.57	18.71
ϕ	-1.86	-1.86	-1.91	-1.84	-1.95	-1.95	-1.89	-2.03	-1.91	-1.98
a	0.30	0.20	0.40	0.30	0.20	0.20	0.50	0.40	0.30	0.20
Panel D: Model$_4$										
Intercept	2.45 ***	0.52 *	1.13	-4.05	1.51	0.81	2.21	2.49	1.08	0.88
ρ_1	0.63 ***	2.27 ***	0.80 ***	0.18 ***	1.06 ***	0.99 ***	0.42 ***	1.11 ***	0.70 ***	0.82 ***
ρ_2	0.06 ***	-2.02 ***	0.60 ***	0.02 ***	-0.28 ***	-0.22 ***	0.15 ***	-0.18 ***	0.18 ***	0.42 ***
ρ_3	0.30 ***	0.74 ***	-0.41 ***	0.19 ***	0.20 ***	0.22 ***	0.36 ***	0.04 ***	0.09 ***	-0.24 ***
ω_1	0.43 ***	-1.29 ***	0.17 ***	0.73 ***	0.03 ***	-0.04 ***	0.56 ***	-0.15 ***	0.25 ***	0.22 ***
ω_2	-0.06 ***	-1.21 ***	-0.26 ***	0.26 ***	0.01 ***	-0.09 ***	-0.20 ***	0.00 ***	-0.34 ***	-0.15 ***
ω_3	-0.24 ***	-0.47 ***	0.17 ***	-0.01 ***	-0.04 ***	-0.19 ***	-0.19 ***	0.12 ***	-0.11 ***	0.23 ***
Panel E: Model$_5$										
ρ_1	0.91 ***	0.85 ***	0.83 ***	0.83 ***	0.82 ***	0.87 ***	0.81 ***	0.78 ***	0.78 ***	0.89 ***
ρ_2	-0.36 ***	-0.33 ***	-0.31 ***	-0.27 ***	-0.30 ***	-0.16 ***	-0.34 ***	-0.28 ***	-0.16 ***	-0.26 ***
ρ_3	0.07 ***	0.10 ***	0.10 ***	0.04 ***	0.07 ***	0.00 ***	0.10 ***	0.08 ***	0.03 ***	0.03 ***
β	0.02 ***	0.08 ***	0.04 ***	0.03 ***	0.04 ***	0.10 ***	0.06 ***	0.10 ***	0.10 ***	0.08 ***
ω_0	7.80 ***	7.91 ***	6.04 ***	6.10 ***	7.16 ***	6.08 ***	7.32 ***	6.54 ***	7.34 ***	7.03 ***
ω_1	7.44 ***	10.22 ***	9.41 ***	9.55 ***	5.23 ***	6.54 ***	7.32 ***	3.22 ***	8.54 ***	4.77 ***
ϕ	-0.24 ***	-0.17 ***	-0.23 ***	-0.16 ***	-0.27 ***	-0.19 ***	-0.11 ***	-0.27 ***	-0.22 ***	0.23 ***

Legend:

- ***** denotes significant at 5% LOS (Level of Significance)
- ****** denotes significant at 1% LOS
- ******* denotes significant at 0.1% LOS

Table 3. Parameter Estimates (Continued)

	ATL	CVG	DFW	DSM	LAS	LGA	ORD	PDX	PHL	TUS
Panel F: Model₆										
β_0	14.98***	13.04***	22.12***	11.09***	15.69***	12.47***	10.08***	13.94***	15.13***	19.72***
β_1	0.0002	0.0004*	0.00008	0.0002*	0.0003***	0.0001**	0.0003	0.0002	0.0003***	0.0002***
$\delta_{c,1}$	-4.47***	-5.99***	-7.61***	-7.01***	-5.53***	-5.66***	-6.20***	-3.96***	-6.66***	-5.56***
$\delta_{c,2}$	-0.22**	-0.05	0.002		-0.17**	-0.15	-0.13	-0.40***	-0.13	-0.21***
$\delta_{c,3}$	-0.12	0.08	-0.31*		-0.02	-0.18	-0.44***	0.02	-0.22*	-0.19***
$\delta_{c,4}$	-0.10	-0.01	-0.25		-0.01	-0.03	-0.28*	0.04		0.10
$\delta_{c,5}$		0.03			0.03	-0.06	0.03			
$\delta_{s,1}$	-0.82***	-1.40***	-2.25***	-1.32***	-1.37***	-1.47***	-1.18***	-1.35***	-2.10***	-1.53***
$\delta_{s,2}$	0.22*	0.04	0.59***		1.06***	-0.33**	0.19	0.59	-0.06	0.60***
$\delta_{s,3}$	-0.06	-0.25	0.06		-0.15*	-0.16	0.09	-0.02	-0.31**	0.29***
$\delta_{s,4}$	-0.14	-0.04	0.18		0.26***	0.04	-0.06	0.12		-0.07
$\delta_{s,5}$		-0.02			-0.07	-0.13	-0.18			
ρ_1	0.96***	0.89***	0.82***	0.87***	0.97***	0.83***	0.87***	0.87***	0.83***	0.91***
ρ_2	-0.35***	-0.31***	-0.24***	-0.28***	-0.29***	-0.28***	-0.24***	-0.12***	-0.26***	-0.23***
ρ_3	0.10***	0.16***	0.08***	0.15***	0.10***	0.14***	0.12***	-0.04*	0.17***	0.04**
ρ_5	-0.02	-0.05***	0.01	0.02	-0.01	-0.06***	-0.06***	0.02	-0.06***	-0.03
ρ_6	0.02	0.04*	-0.02	-0.01		0.12***	0.03	0.03	0.07***	-0.01
ρ_7	-0.01	0.01	0.03	-0.03*		-0.06***	0.03	-0.03	-0.03	0.02
ρ_8	0.01	0.05**	-0.01	0.03*		0.06***	0.02	0.05**	0.04**	0.003
ρ_9	0.01	-0.08***	-0.03**	-0.01		0.001	-0.04*	-0.04*	0.01	-0.03
ρ_{10}	0.00	0.02		0.03		-0.04*	0.01	0.04*	-0.03	0.02
ρ_{11}	-0.01	-0.05**		-0.05**		0.04*	-0.05**	-0.01	0.01	0.01
ρ_{12}	0.01	0.05**		0.02		-0.04*	0.06**	0.02	-0.03*	-0.02
ρ_{13}	0.02	0.03		0.05**		0.03	-0.02	-0.04*	0.04*	0.02
ρ_{14}	-0.01	-0.07***		-0.11***		-0.03	0.01	0.03	-0.05**	0.002
ρ_{15}	-0.01	0.08***		0.07***		0.01	-0.01	0.05**	0.01	-0.04*
ρ_{16}	0.02	-0.07***		-0.03		-0.01	-0.01	-0.09***	-0.04*	0.06***
ρ_{17}	0.003	0.06***		0.03		0.001	0.02	0.01	0.001	-0.04*
ρ_{18}	0.01	0.02		0.003		0.03	-0.01	0.01	0.02	0.01
ρ_{19}	-0.02	0.00		0.04**		0.001	0.02	-0.04*	0.02	-0.01
ρ_{20}	0.01	-0.01				0.01	0.06***	0.02	-0.02	0.002
ρ_{21}	-0.01	0.004				-0.04*	-0.06***		0.01	0.03
ρ_{22}	0.03					0.06**	0.03*		0.06***	-0.05**
ρ_{23}	-0.02					0.02			-0.01	0.03
ρ_{24}	0.02					-0.02			0.01	0.02
ρ_{25}	0.001					0.001			-0.04*	-0.01

Legend: * denotes significant at 5% LOS (Level of Significance)
** denotes significant at 1% LOS
*** denotes significant at 0.1% LOS

Table 4. Descriptive Statistics for In-Sample and Out-of-Sample Forecasts

	AccHDD			AccCDD			Avg T		
In-Sample	**Mean**	**Median**	**StdDev**	**Mean**	**Median**	**StdDev**	**Mean**	**Median**	**StdDev**
Actual	560.35	506.50	346.31	235.29	176.50	215.36	58.32	59.50	18.53
$Model_1$	539.97	529.21	330.10	228.80	175.16	205.27	58.37	59.11	18.20
$Model_2$	564.75	545.12	347.02	229.99	160.80	216.01	57.94	58.29	18.34
$Model_3$	560.75	515.38	316.13	233.07	166.23	231.97	58.54	58.95	18.18
$Model_4$	555.49	507.18	352.05	226.89	164.51	220.46	58.32	59.15	17.88
$Model_5$	555.08	506.57	352.47	226.58	164.09	220.55	58.32	59.16	17.85
$Model_6$	578.15	534.09	351.97	221.79	156.87	217.18	57.76	58.74	18.16
	AccHDD			**AccCDD**			**Avg T**		
Out-of-Sample	**Mean**	**Median**	**StdDev**	**Mean**	**Median**	**StdDev**	**Mean**	**Median**	**StdDev**
Actual	527.93	469.00	323.33	252.76	198.00	227.38	59.30	59.50	19.62
$Model_1$	544.14	491.14	323.18	236.15	193.70	210.42	58.46	58.87	18.37
$Model_2$	549.55	478.73	370.55	230.18	180.94	216.23	58.38	59.80	18.49
$Model_3$	567.33	521.89	320.03	232.63	170.88	231.27	58.38	58.77	18.44
$Model_4$	550.82	517.47	347.59	222.02	164.47	226.98	58.42	58.95	17.24
$Model_5$	553.43	516.85	345.27	217.83	161.14	220.34	58.29	58.83	17.12
$Model_6$	540.88	520.79	336.49	224.08	170.34	223.36	58.69	59.17	16.89

Legend: **Avg T = Daily Average Temperature**
AccHDD = Monthly-accumulated HDD Index
AccCDD = Monthly-accumulated CDD Index

Table 5. In-Sample and Out-of-Sample Forecasting Error (Relative Error)

In-Sample	AccHDD			AccCDD			Avg T		
	Mean	Median	t-value	Mean	Median	t-value	Mean	Median	t-value
$Model_1$	4.21%	4.42%	7.44	1.87%	0.67%	2.79	0.27%	0.73%	3.83
$Model_2$	1.21%	2.16%	1.75	3.16%	1.02%	4.08	1.35%	1.50%	20.54
$Model_3$	1.93%	2.20%	3.47	2.83%	0.31%	4.28	0.28%	0.75%	5.02
$Model_4$	1.08%	0.20%	8.81	2.37%	0.21%	10.06	0.13%	0.61%	3.66
$Model_5$	1.21%	0.20%	8.50	2.44%	0.24%	10.34	0.12%	0.63%	3.38
$Model_6$	-3.14%	-2.95%	-16.57	4.60%	2.21%	13.65	1.19%	1.40%	32.74
Out-of-Sample	**AccHDD**			**AccCDD**			**Avg T**		
	Mean	**Median**	**t-value**	**Mean**	**Median**	**t-value**	**Mean**	**Median**	**t-value**
$Model_1$	-2.74%	-1.50%	-1.66	0.96%	0.00%	0.43	1.12%	1.47%	7.22
$Model_2$	-1.81%	-0.25%	-0.94	3.29%	0.61%	1.51	1.25%	1.59%	8.61
$Model_3$	-2.12%	-0.47%	-1.40	2.59%	0.00%	1.42	1.40%	1.58%	11.65
$Model_4$	5.97%	9.37%	2.91	-2.38%	0.00%	-1.13	1.42%	1.42%	8.20
$Model_5$	5.99%	9.38%	2.92	-2.50%	0.00%	-1.18	1.61%	1.44%	9.42
$Model_6$	-1.78%	-0.81%	-1.23	2.24%	0.00%	1.26	0.92%	1.31%	8.99

Legend:
Avg T = Daily Average Temperature
AccHDD = Monthly-accumulated HDD Index
AccCDD = Monthly-accumulated CDD Index

Table 6. Best Model by Cities

	In-Sample			Out-of-Sample		
	AccHDD	**AccCDD**	**Avg T**	**AccHDD**	**AccCDD**	**Avg T**
ATL	Model 5	Model 5	Model 4	Model1	Model1	Model6
CVG	Model 4	Model 4	Model 5	Model4	Model1	Model6
DFW	Model 5	Model 5	Model 6	Model1	Model2	Model6
DSM	Model 1	Model 5	Model 6	Model1	Model1	Model6
LAS	Model 4	Model 4	Model 6	Model1	Model1	Model6
LGA	Model 4	Model 4	Model 5	Model6	Model1	Model6
ORD	Model 4	Model 4	Model 6	Model6	Model1	Model6
PDX	Model 4	Model 6	Model 6	Model6	Model3	Model6
PHL	Model 5	Model 4	Model 5	Model6	Model1	Model3
TUS	Model 5	Model 4	Model 5	Model2	Model1	Model6

Legend: **Avg T = Daily Average Temperature**
AccHDD = Monthly-accumulated HDD Index
AccCDD = Monthly-accumulated CDD Index

Figure 1. Seasonality in Daily Average Temperature for Atlanta and Chicago.

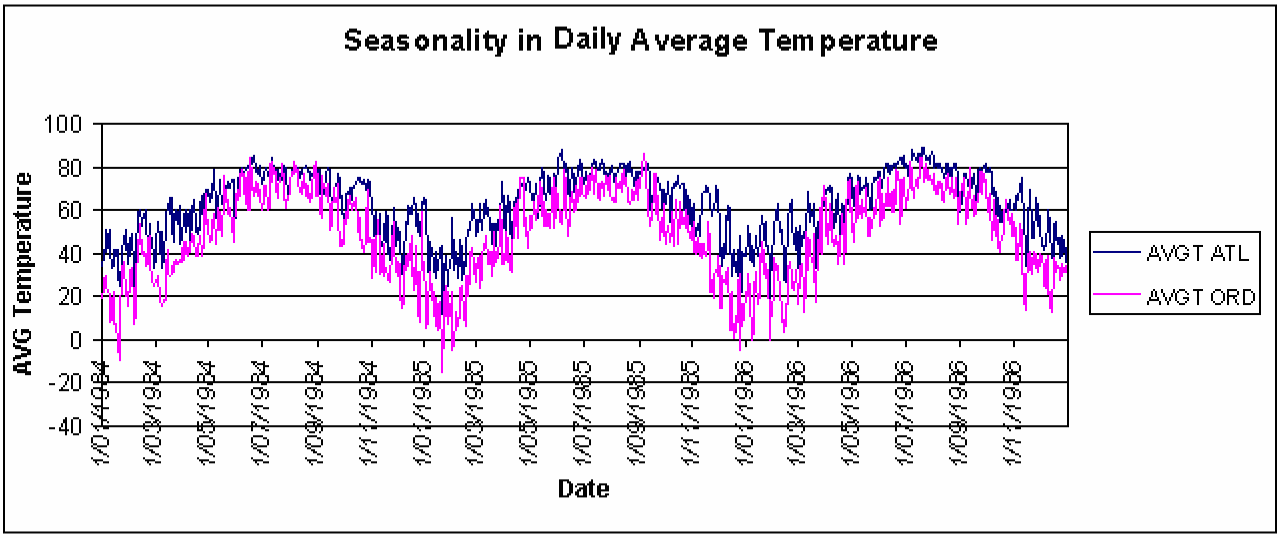

Figure 2. In-Sample UAPE

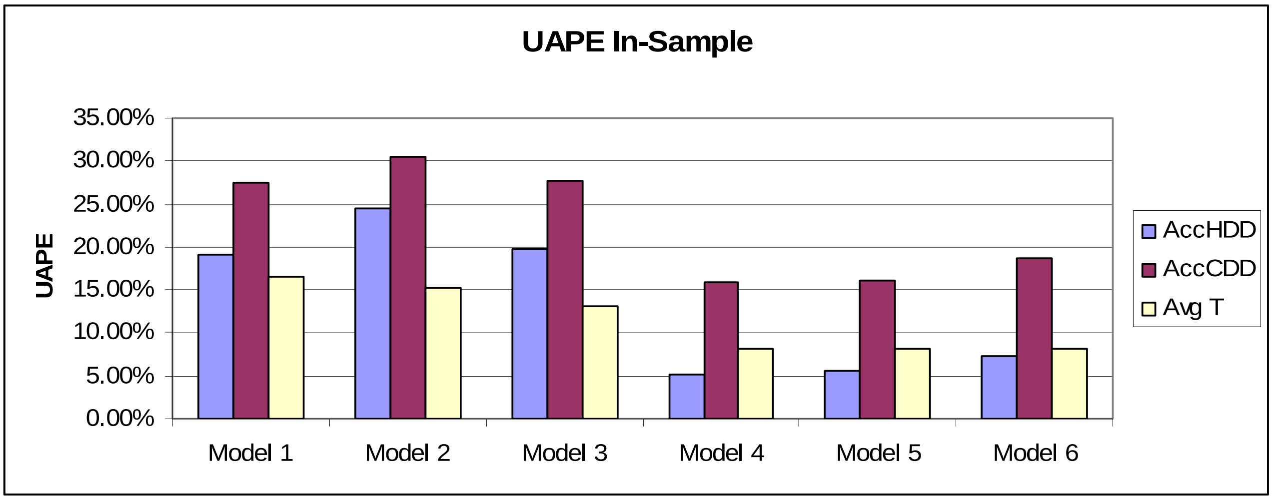

Figure 3. Out-of-Sample UAPE

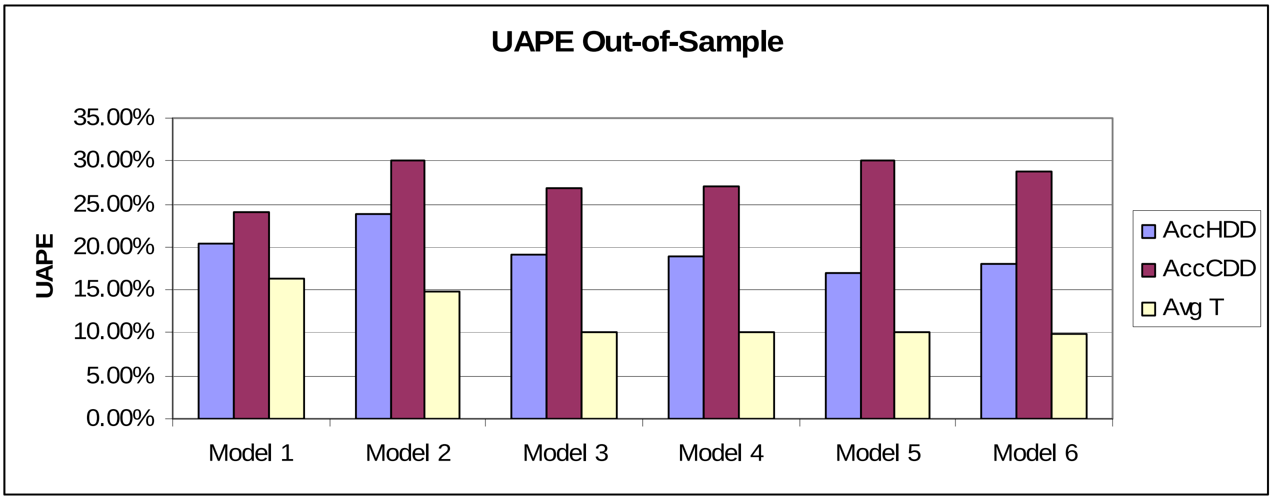

Figure 4. UAPE In-Sample by Month (Monthly-accumulated HDD)

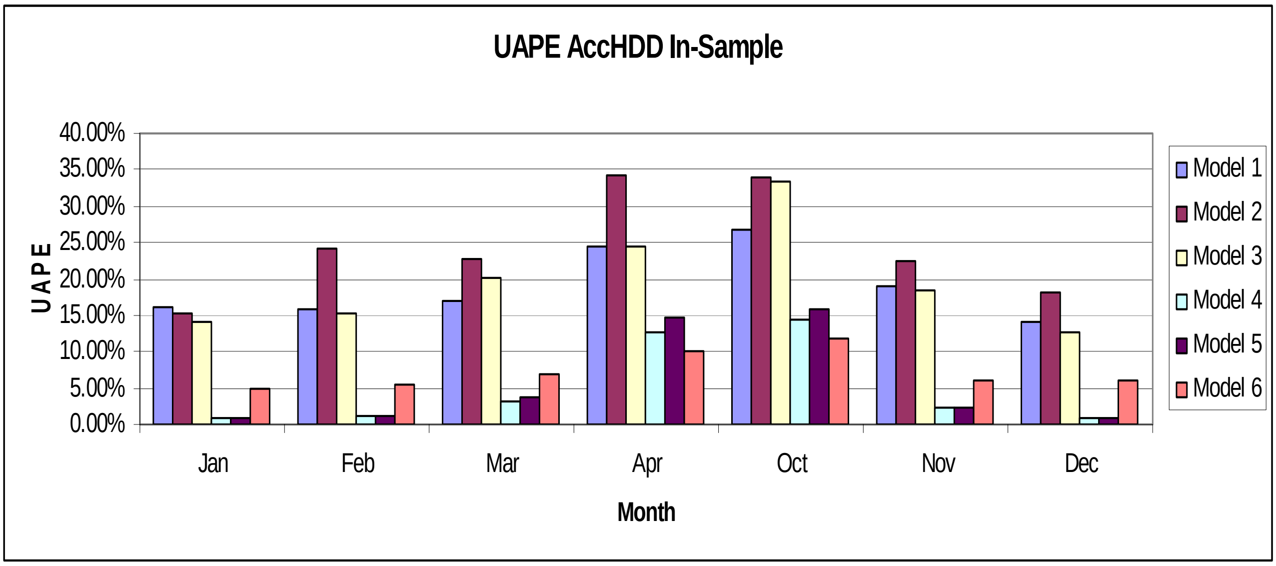

Figure 5. **UAPE In-Sample by Month (Monthly-accumulated CDD)**

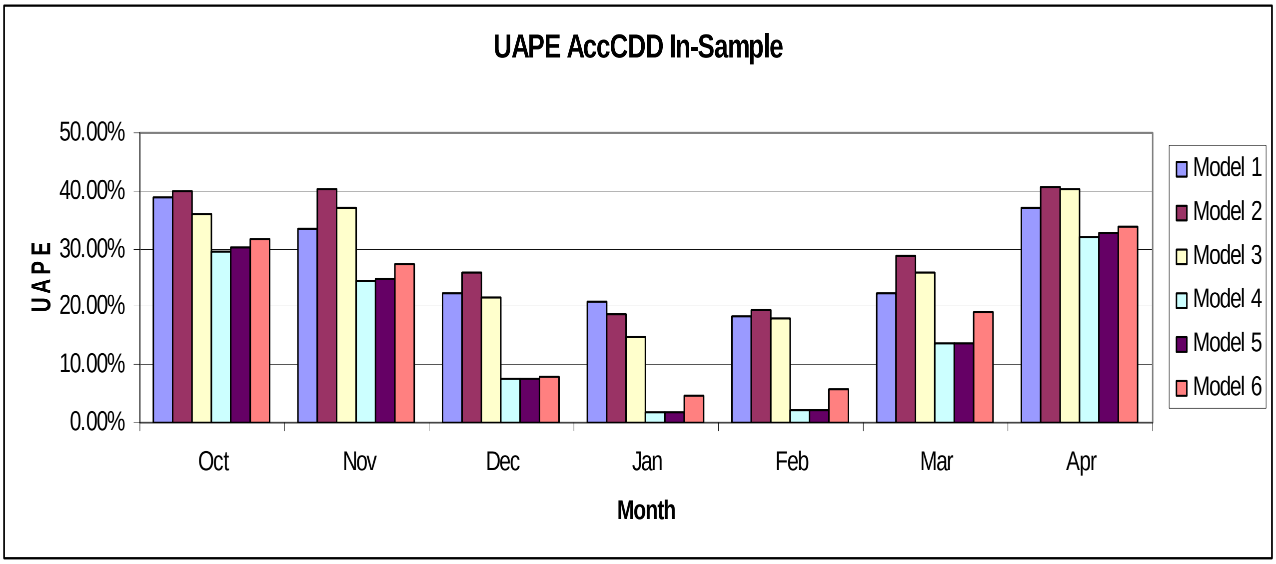

Figure 6. Out-of-Sample UAPE by Month (Monthly-accumulated HDD)

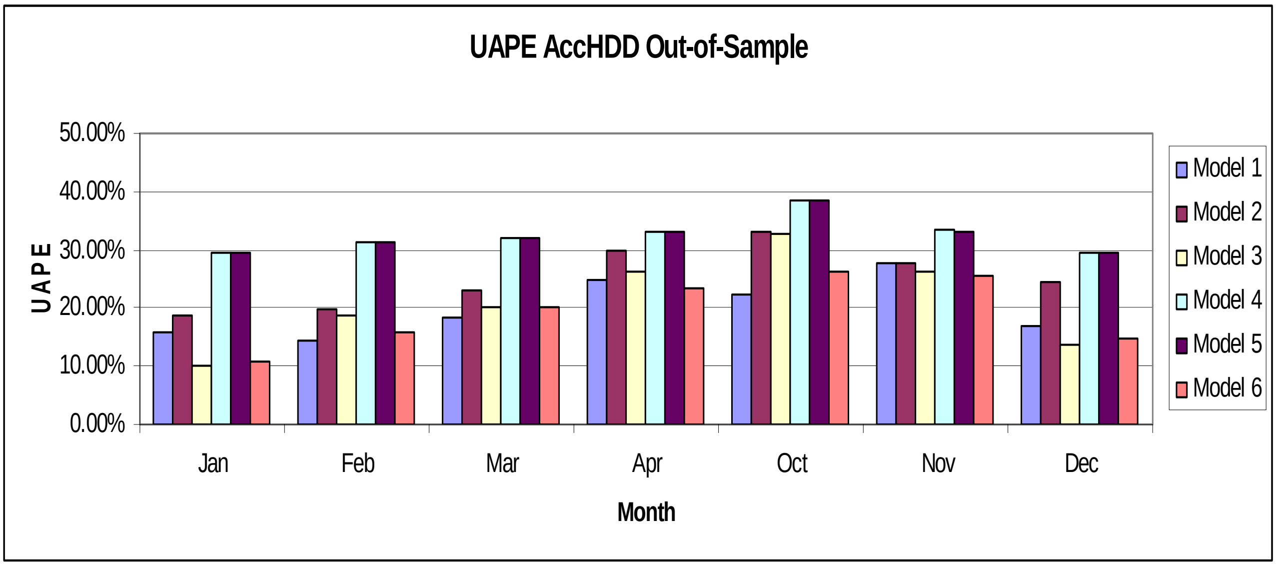

Figure 7. Out-of-Sample UAPE by Month (Monthly-accumulated CDD)

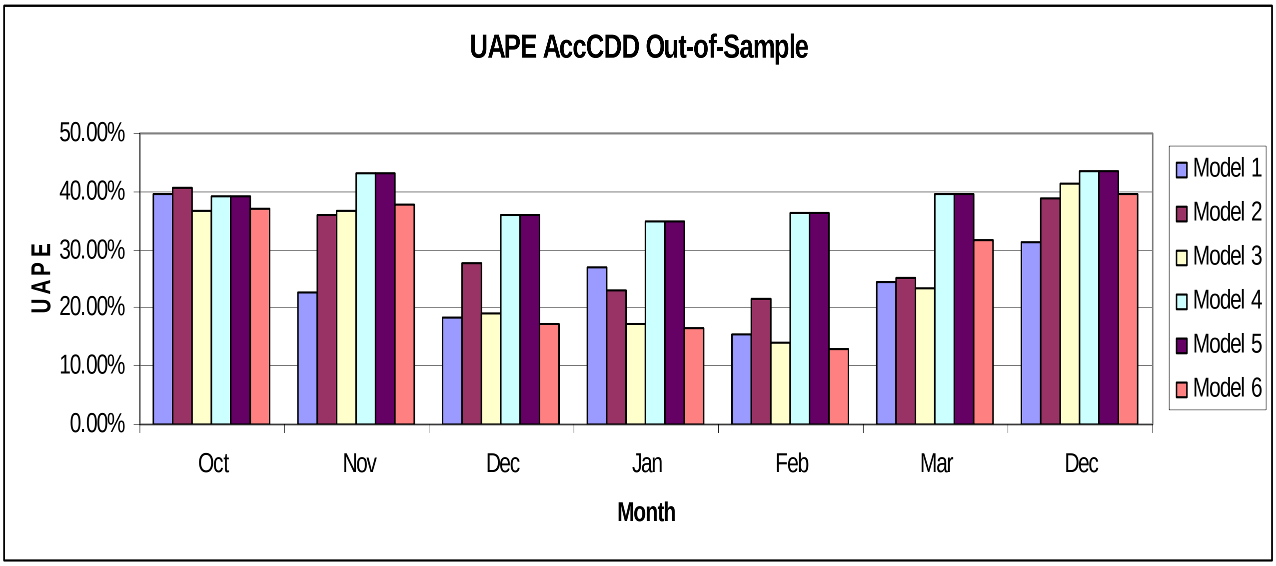

Figure 8. Out-of-Sample Step Ahead Forecast (AccHDD)

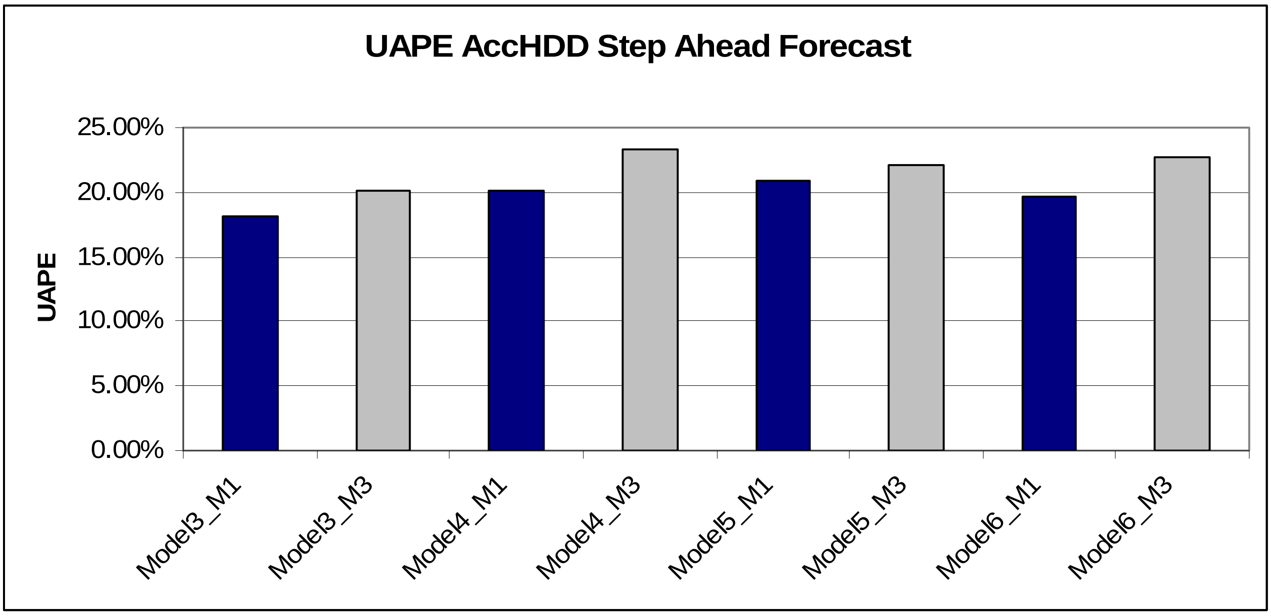

Figure 9. Out-of-Sample Step Ahead Forecast (AccCDD)

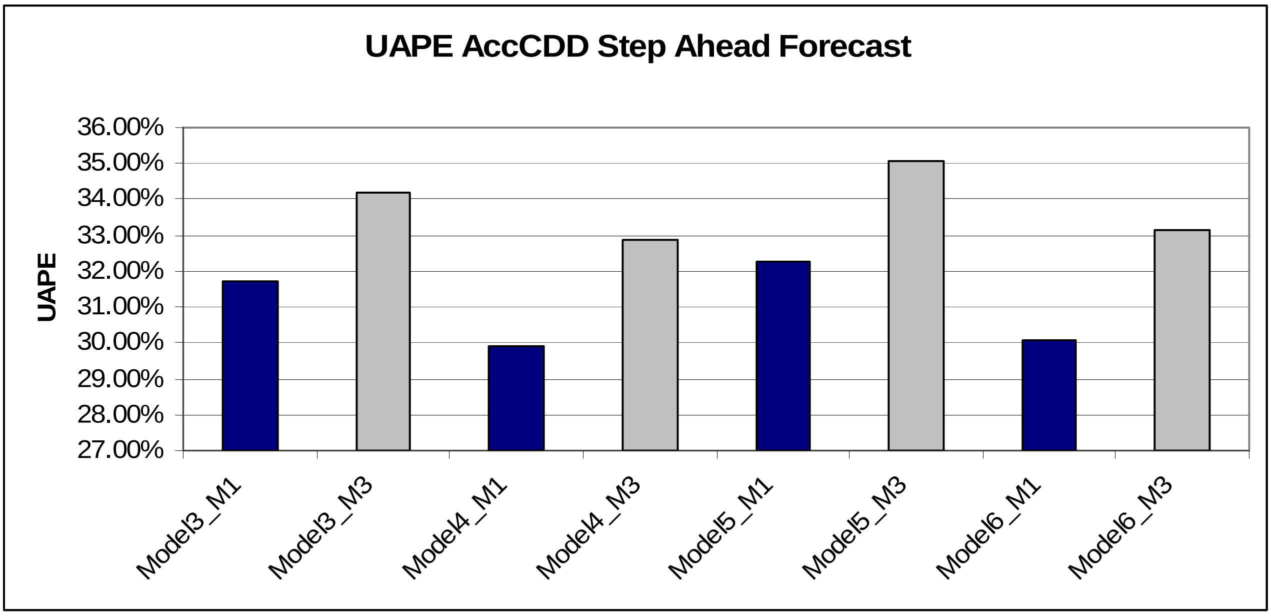

Figure 10. Out-of-Sample Step Ahead Forecast (Daily Average Temperature)

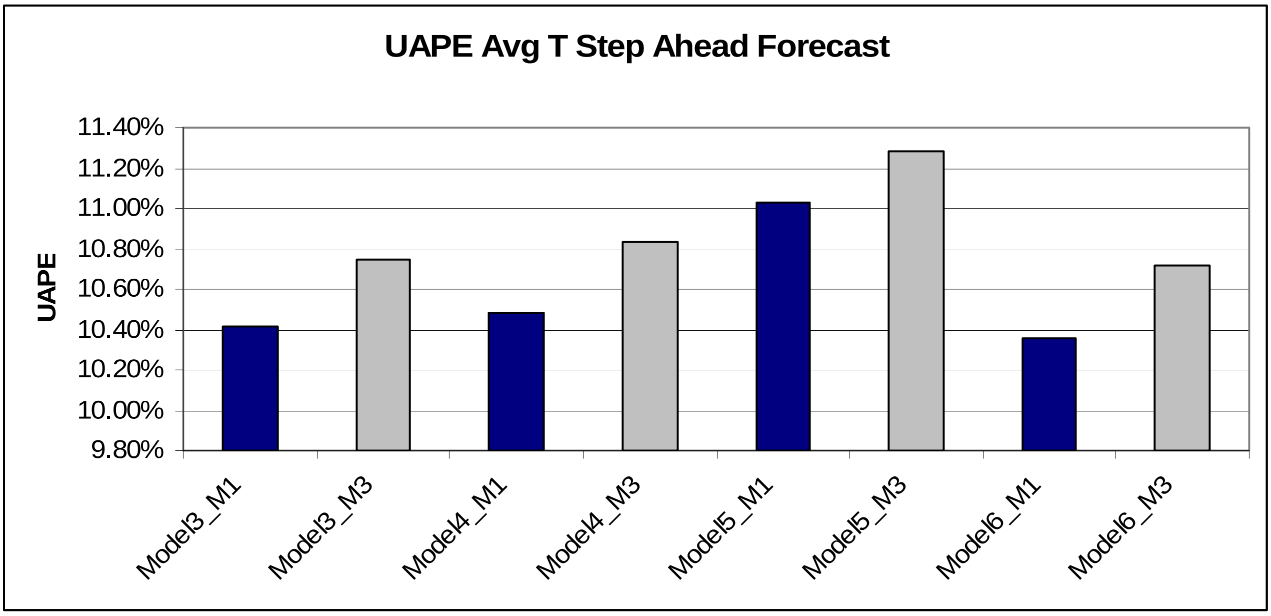

Figure 11. Out-of-Sample Comparison against Contract Price (Monthly-Accumulated HDD Contracts)

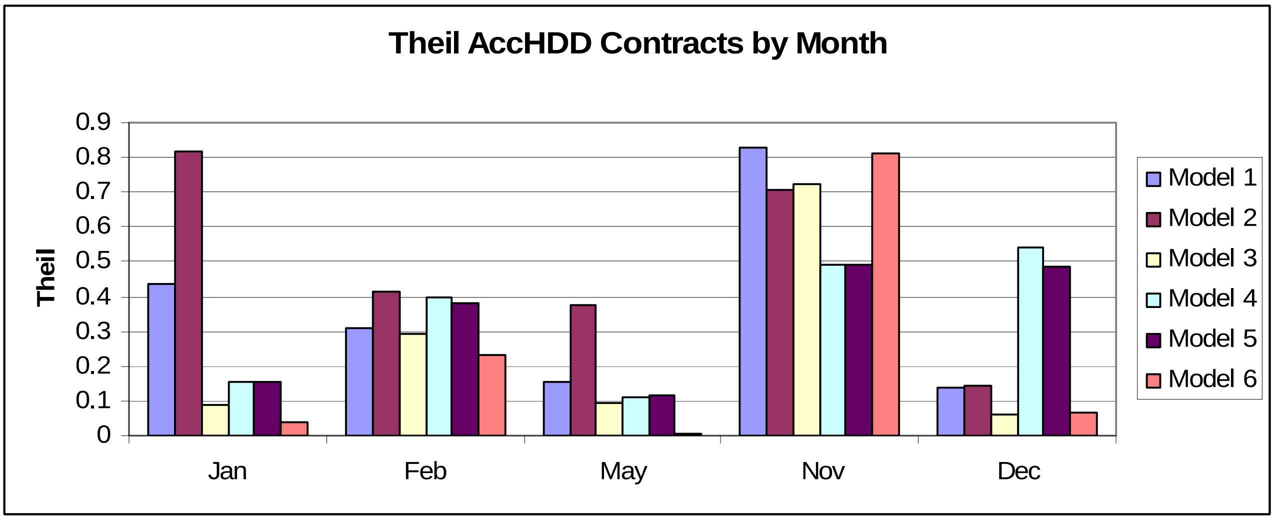

Figure 12. Out-of-Sample Comparison against Contract Price (Monthly-Accumulated CDD Contracts)

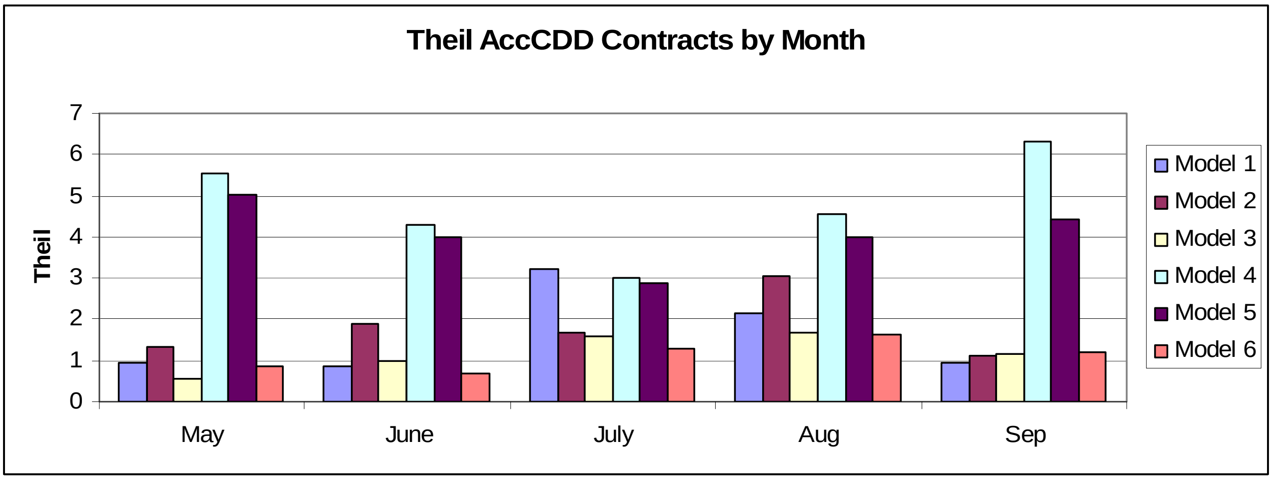

Figure 13. Out-of-Sample Comparison against Contract Price by City (Monthly-Accumulated HDD Contracts)

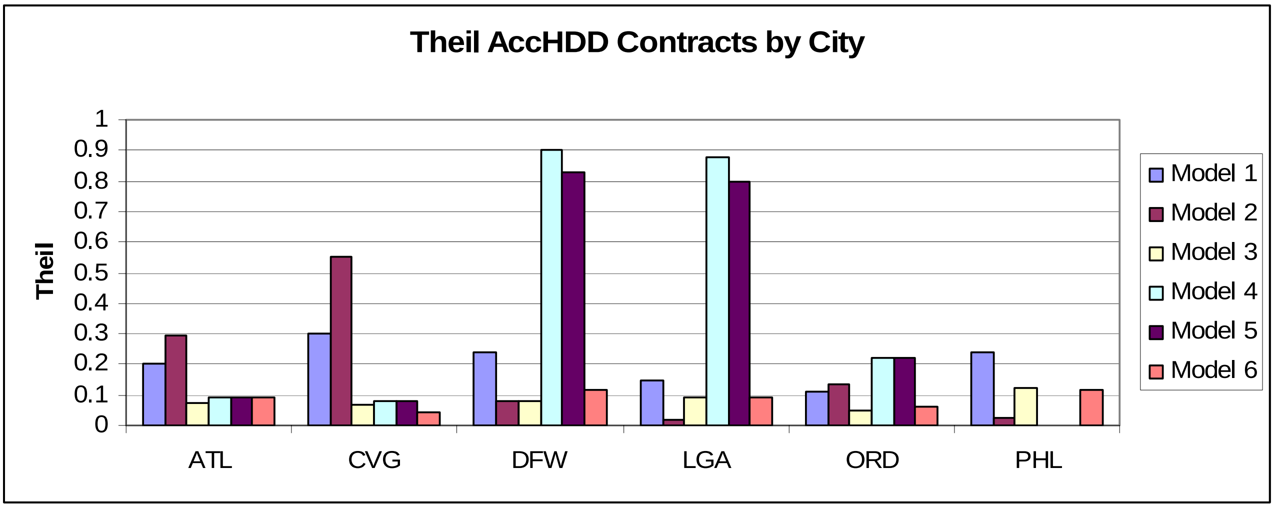

Figure 14. Out-of-Sample Comparison against Contract Price by City (Monthly-Accumulated CDD Contracts)