UNIVERSITY OF PIRAEUS

DEPARTMENT OF BANKING AND FINANCIAL MANAGEMENT

GRADUATE PROGRAM

Forecasting Temperature with a View to Weather Derivatives

Garen Papazian

Supervisor: George Skiadopoulos, PhD

Piraeus, September 2008

Acknowledgments

First and foremost, I wish to thank Professor George Skiadopoulos for accepting to be the supervisor of my thesis and for his inspiration, constant assistance, and valuable comments. He has always been patient and has always found time to discuss problems and ideas with me. The interest that he showed in the project was beyond expectations.

I also wish to remember Eirini Konstantinidi and Professor Fred Benth for their fair comments and suggestions.

Thanks to my friends Panagiotis and Lambros for their friendship and for the beautiful time spent together during the years of our postgraduate studies. Hope you all enjoyed yourselves as much as I did.

Especially, I would like to thank my friends John, Kostas, Antonis, Leonidas, Vlassis, Vasia and Krikor for their unremitting friendship from my school days until now. Without their support and faith on me I would not have reached this level.

Finally, I would like to thank Elsa and my family for their love, support and patience over the last most difficult years of my life. Hope I have not let you down.

Table of Contents

Appendices

List of Figures

List of Tables

Abstract

With the rapid development of the weather derivatives market, researchers have proposed various models for the representation of the behavior of weather variables. The objective of this study is to obtain robust results between proposed temperature forecasting models - Campbell and Diebold (2005) and Benth and Saltyte-Benth (2005) - used as the underlying processes for valuing weather derivatives contracts, based on their performance in different locations and in different horizons. Furthermore, we propose and implement new ways - a first order autoregressive process, the principal component analysis (PCA) and the combining forecasting approach - of modeling and forecasting the temperature behavior. The statistical significance of these forecasts is examined. Finally, we attempt to dynamically forecast the two most traded weather indices in the CME market, the cumulative heating degree day (CumHDD) and the cumulative cooling degree day (CumCDD) indices. The statistical significance of these forecasts is examined using as benchmark the values produced by the "burn analysis" methodology.

After examining the different forecasting models, we found that no one model was able to consistently outperform the others. The most appropriate forecasting model varied between cities and horizons. Thus, a general model that adequately explains the temperature is absent and probably not possible. However, we tried the PCA for forecasting purposes and we found that it could be a very powerful method specifically for the temperature of the U.S.A. The combining forecasting approach performed best for the temperature of Europe. The index forecasting exercise showed that the CumHDD index is more difficult to be forecasted than the CumCDD index. Quite interestingly, the European weather indices can be forecasted more adequately than those of the U.S. since the accuracy measures are significantly lower. As far as regards the second objective of this exercise, we concluded that the forecasts calculated by the naïve "burn analysis" methodology should not be rejected a priori in every case.

1. Introduction

Almost every business activity is exposed to adverse weather conditions. The Department of Commerce of the United States conducted a research about the weather risk affecting the U.S economy and found that about $1 trillion of the $7 trillion US economy is directly exposed to it. Moreover, more recent researches have shown that about 70% of companies encounter risks relevant with the weather. However, when we are talking about adverse weather conditions we do not only consider low probability, highly catastrophic events. Business can be affected easily by small disturbances of weather dynamics that the insurance industry cannot satisfy.

Lately, a new class of financial instruments, weather derivatives, has been introduced to enable businesses to manage their *volumetric risk* - the uncertainty on their volume exposure regarding both the number of customers and their demanded goods volume - resulting from adverse weather patterns.[1] A perfect hedge requires hedging both price risk, by way of standard commodity derivatives, and volume risk, by way of weather derivatives. Hence, the objective of these new instruments is to hedge volume risk that results from an alteration in the demand of commodities due to unexpected changes in weather.

Traditional contingent claims have payoffs that depend upon the price of some fundamentals. Likewise, a weather derivative has as its underlying asset a weather variable. Financial contingent claims are priced by no-arbitrage arguments, such as Black-Scholes pricing model, based on the concept of continuous hedging. The main assumption of this model is that the underlying asset of the contract can be continuously traded. However, in the case of weather derivatives this hypothesis is violated and no riskless portfolio can be constructed. Until now the pricing of weather contracts is one of the hardest problems still to be solved.

Two alternatives pricing methodologies can be followed to obtain the "fair value" of this new class of contingent claims.[2] The first approach is called "burn ana-

[1] For a brief description about the weather derivatives market, we refer the reader to Appendix I.

[2] See Jewson and Brix (2005), "Weather Derivative Valuation", for an analytical review of the pricing methodologies.

lysis" and it is mostly adopted by the insurance industry. The burn analysis is based simply on the idea of evaluating how a contract would have performed every year in the past and then calculating the mean of the historical values. Although, there are cases when this method is quite inaccurate, - due to its simplicity assuming that weather variable would behave as it did in the past - is a good first step in pricing almost any type of contract.

The second is called "weather modeling" and is more complicated, since it aims to model and forecast the weather behavior directly (see Alaton, Djehiche, Stillberger, 2002, Benth and Saltyte-Benth, 2005, 2007). Yet, the evolution of weather differs significantly from that of securities prices. Thus, it is of crucial importance to carefully validate the specified model before putting into practical use for pricing weather derivatives.

This study concentrates on the most widely traded weather derivatives, temperature-based derivatives. In 2006 PriceWaterhouseCoopers conducted a survey showing that more than 95% of the notional values of weather contracts - including both over the counter (OTC) & CME market standardized contracts - are temperature-based contracts. This is attributed to the high participation of the energy industry in the weather market due to the high correlation between temperature and energy demand.

The pay-off functions of temperature-based contracts are financially settled using as input values the measured values of various temperature indices. The most commonly used indices are degree day indices, average temperature indices, cumulative average temperature indices and event indices. Degree day indices originated in the energy industry and are designed to correlate well with the demand for heating and cooling. In winter, heating degree days (HDDs) are used to measure the demand for energy used for heating (the lower the temperature, the higher the HDD). In summer, cooling degree days (CDDs) are used to measure the demand for energy used for cooling (the higher the temperature, the higher the CDD).

The last decades in parallel with the rapid development of the weather derivatives market, researchers have proposed temperature forecasting models that can be integrated into the pricing framework as the underlying processes and at the same time provided accurate estimates and forecasts, either in continuous or in time series approaches. Cao and Wei (2004) first detrend and deseasonalize the temperature series by subtracting the historical mean. Then they model the daily temperature residuals with an autoregressive process with a seasonal conditional variance. Campbell and Diebold (2005) extended the autoregressive model of the previous paper. They propose a high parameter model that possesses seasonal components by means of a low order Fourier series and autoregressive components selected by the Akaike and Schwarz criteria. The conditional variance equation contains generalized autoregressive conditional heteroscedasticity (GARCH) dynamics (Engle 1982; Bollerslev 1986) with a seasonal component by means of a Fourier series.

Dischel (1998) employed a continuous process such as those used for the explanation of the short term interest rate incorporating a parameter that explains the salient feature of seasonality. The criticism to the paper of Dischel (1998) was done by Dornier and Querel (2000), who did not approve the use of the mean reverting parameter to illustrate the trend and the seasonality. They argued that the model did not revert to the long run mean. They showed also that the answer to the problem is the addition of a parameter explaining the changes of seasonal variation. Alaton, Djehiche and Stillberger (2002) improved the work of Dischel (1998) by incorporating the suggestion of Dornier and Querel (2000), while in the same time modeled the mean seasonality with a sine wave function. The standard deviation of temperature is modeled by a piecewise function that varies across different months and remains constant in each month. Torro, Meneu and Valor (2003) modeled air temperature behavior in Spain with combining methods used for the modeling of short term interest rates and a generalized autoregressive conditional heteroscedastic (GARCH) time series approach. They observed that the quadratic variation is well explained by the GARCH model. However, they did not incorporate the adjusting factor proposed by Dornier and Querel (2000) and hence their model did not produce a consistent mean reversion to long run mean.

Ending the scarce literature about the modeling of the temperature we refer to the papers of Benth and Saltyte-Benth (2005, 2007). The modeling of daily temperature variations is done with a mean reverting Ornstein-Uhlenbeck process with a seasonal mean and volatility. They found that the proposed dynamics fitted quite successfully the Norwegian temperature data. Interestingly, the researchers accomplish also to derive straightforward prices for temperature derivatives.

This paper makes at least three contributions to the ongoing discussion about temperature forecasting with a view to weather derivatives. First, it employs an extensive set of U.S. and European temperature data; the dataset contains daily temperature observations from ten U.S. and five European weather stations ranging from January 1, 1973 to December 31, 2007. We will attempt to obtain robust results between prior proposed models - Campbell and Diebold (2005) and Benth and Saltyte-Benth (2005) - based on their performance in different locations. Second, we propose and implement new ways - a first order autoregressive process, the principal component analysis (PCA) and the combining forecasting approach - of modeling and forecasting the temperature behavior. Point forecasts are formed and evaluated in short and longer horizons. Third, we enhance our analysis by attempting to dynamically forecast the two most traded weather indices in the standardized CME market, the monthly cumulative heating degree day (CumHDD) and cumulative cooling degree day (CumCDD) indices (see Stevenson and Oetomo, 2005, for a similar approach). In specific, we use two forecasting models - Campbell and Diebold (2005) and Benth and Saltyte-Benth (2005). This exercise would be valuable for taking positions in the weather derivatives market.

To fix ideas, the possible presence of an inexpensive, simple and extensible time series model it may prove useful for weather derivative's pricing purposes in the newly established market of Chicago Mercantile Exchange. Weather forecasting is of crucial importance since any firm or country exposed to weather risk on either its output (revenue) side or the input (cost) side wants to know how the weather would behave. In addition, the potential adequacy of a time series model would be also helpful for companies to hedge their weather risk. Finally, as the weather market has evolved to the point where hedge funds and other non-commercial traders trade CME weather products to absorb risk in exchange for possible profit on weather variations the necessity for adequate forecasting models expands.

The remainder of the paper is structured as follows. In the next section, the data sets are described. In section 3 we present the models to be used for forecasting. The out-of-sample performance of the generated forecasts is evaluated in sections 4.

In section 5 we present the index forecasting procedure along with the respective results. The last section concludes.

2. The data set

Our dataset contains daily temperature observations measured in Fahrenheit degrees, for two groups of measuring stations.[3] Specifically, the first group consists of ten cities from the United States and the second group consists of five cities from Europe. The choice of our sample is based on locations, where weather derivatives contracts are traded and are regarded as global financial markets. [4] We obtained the data from Earth Satellite (Earth Sat) corporation; they are precisely those used to settle temperature-related weather derivative products on the CME, a fact important for the appreciation of our data. The underlying data source is the National Climatic Data Center (NCDC), a division of the National Oceanographic and Atmospheric Administration. Each of the measuring stations of our groups supplies its data to the NCDC, and those data are in turn collected by Earth Sat.

The data comprise actually measured daily maximum and minimum temperature (measured in Fahrenheit). Table 1 shows the corresponding measuring stations for U.S.A and Europe in panels A and B, respectively. The data sample ranges from January 1, 1973 to December 31, 2007. We construct the daily average temperature series by averaging the daily maximum and minimum temperature. Following Campbell and Diebold (2005) and Taylor and Buizza (2006), we have discarded the 29th of February in leap years to maintain 365 days per year. So we have a total of 12,775 daily observations. The dataset contains an insignificant fraction of missing observations for our data in the groups of U.S.A. and Europe, due to the measuring stations failure to report to the NCDC. These missing observations are attributable to factors such as human error or mechanical failure of the measurement equipment. When missing values are encountered, we fill them with the method proposed by Kosater (2006), who replaced the missing observations by the average of

[3] The formula to convert a Fahrenheit scale temperature into degrees on the Celsius scale is

$$T_C = \frac{5}{9}\left(T_F - 32\right).$$

[4] In U.S.A., the Chicago Mercantile Exchange (CME) lists derivatives on various indices for a set of cities (http://www.cme.com). Our initial sample included the city of Tucson where CME underwrites weather derivatives, but the large gaps in its dataset, induced us to employ the temperature of Los Angeles instead.

temperatures observed one year before and one year after, for the following reasons.[5] The mistakes do not occur in consecutive blocks and their relative appearance is less than 0.05% of the full sample. The alternative method used by Campbell and Diebold (2005), spatial interpolation, requests data from a significant number of nearby reference stations. Yet, the European measuring stations do not have an adequate number of reference stations for reliable results. Therefore, we would like to have a consistent approach in "Gap filling" for our two groups.

The subset from January 1st 1973 to December 31st 2003 will be used for the in-sample evaluation and the remaining data will be used for the out-of-sample one. Figure 1 shows the evolution of temperature corresponding to our full sample.[6] There we observe a significant seasonal behavior. Figure 2 shows the histogram for the data series, where is apparent the non-normality of the data. The distributions are, more or less, bimodal as temperature goes from high levels (summer) to low levels (winter).

Table 2 shows the summary statistics of the daily average temperatures in levels and in first differences. We observe that the temperature is more variable in cold cities, since the standard deviation is higher in cities with lower mean temperatures. For most cities, the temperature series reflects a negative skewness that indicates the higher possibility of having more extremely cold days than extremely hot days. In order to detect the case of stationarity of our series, we perform the Augmented Dickey-Fuller (ADF, 1981) test. The ADF tests show that the temperature series are stationary. Temperature has strong autocorrelation since a cold day is more likely to be followed by a colder and vice versa.[7]

Under the Jarque-Bera statistic we reject the hypothesis of the normal distribution in first differences for each city (see also Benth and Saltyte-Benth, 2007, for similar results for Norway). The average and standard deviation vary from city to city. The series exhibits also negative skewness.

[5] For a thorough analysis of the proposed "Gap Filling" procedures, we refer the reader to Appendix II.
[6] To conserve space we report the figures of two cities from each group.
[7] For parsimony reasons we do not report the autocorrelation coefficients.

3. The forecasting models

Five different approaches to temperature forecasting are examined in this study and one benchmark (random walk). The six models are defined below.

3.1. The Random Walk Model (RW)

The benchmark model is a no change-forecast. This model assumes that daily average temperature follows a random walk and is called the "persistence forecast" in the climatological literature. The RW specification is given by:

$$T_t = T_{t-1} + e_t \tag{3.1}$$

3.2. The Autoregressive Model (AR(1))

Univariate autoregressive models are employed in order to examine whether the evolution of temperature can be forecasted using its previous values. This model employs a constant that is interpreted as the average temperature level and a mean reverting parameter. The AR(1) is employed as a more sophisticated benchmark model, since it employs the salient mean reversion characteristic of temperature. The $AR(1)$ specification is given by:

$$T_t = c_0 + r_{t-1}T_{t-1} + e_t \tag{3.2}$$

3.3. The Benth and Saltyte-Benth Model (BSB)

The third forecasting model is based on the paper of Benth and Saltyte-Benth (2005). This model is enclosed on the group of models (see Dornier and Querel, 2000, Alaton, Djehiche and Stillberger, 2002, Benth and Saltyte-Benth, 2007, for a similar approach) that extend well known financial diffusion processes in order to incorporate the basic features of temperature. Benth and Saltyte-Benth propose the use of an Ornstein-Uhlenbeck process to model the temperature dynamics. The model can be written in a discrete time, as the following additive time series model:

$$T_t = c_0 + c_1 t + \sum_{i=1}^{3}\left(s_{c,i}\cos\left(2p\,i\frac{d(t)}{365}\right) + s_{s,i}\sin\left(2p\,i\frac{d(t)}{365}\right)\right) + u_t \quad (3.3a)$$

[8]

$$\hat{u}_t = r_{t-1}\hat{u}_{t-1} + x_t \qquad x_t = s_t e_t \qquad e_t : (0,1) \quad (3.3b)$$

[9]

The estimation is done in several steps. T_t , the average daily temperature, is first regressed on the seasonal component as in equation (3.3a). c_0 is a constant that is interpreted as the average temperature level, t a deterministic linear trend cos and sin denote the well known trigonometrical functions, $s_{c,p}, s_{c,p}$ control the amplitude, i controls the time frequency of the Fourier series and $d(t)$ is a step function that cycles through 1, 2,…,365 (i.e. 1 denotes January 1, 2 denotes January 2 and so on).[10] Afterwards, the fitted residuals, which are called "temperature anomalies" in the weather derivatives market, are employed in a mean-reverting process with zero constant. In the last step, the volatility function is estimated.

The model is a simple mean-reverting stochastic process, but yet powerful enough to describe the most apparent stylized facts of temperature data like seasonality and mean-reversion. The simplicity of the proposed dynamics allows for explicit calculation of weather derivatives futures prices quoted on the Chicago Mercantile Exchange (CME). They were also able to provide explicit prices for call and put options written on temperature futures typically traded on the CME.

[8] Benth and Saltyte-Benth suggest a cosine function to depict the seasonal component of the time series in their paper. In our case we employ a Fourier series as shown in (3.3a). We employed both equations and we found that our equation produces better fit than that suggested by Benth and Saltyte-Benth. Our choice of order 3 in the Fourier series is based both on the R^2 and on the AIC. For this and for parsimony purposes in our study we employ the traditional Fourier series as shown in equation (3.3a). In advance, we employ a linear trend which we find statistically significant in most cases in our sample.

[9] The variance of residuals is estimated in the following way. First, the observed residuals are organized into 365 groups, one for each day of the year. Finding the average of the squared residuals in each group, we obtain an estimate for the expected daily squared residual. In this way, an estimate of s_t^2 is obtained . Note that s_t^2 is assumed to be a periodic function such that $s_t = s_{t+k\cdot 365}$ for $t = 1,\ldots,365$ and $k = 1,2,3,\ldots$.

[10] For a more precise description of the Fourier series we refer the reader to Appendix III.

3.4. The Campbell-Diebold model (CD)

Campbell and Diebold (2005) propose a high order autoregressive process with seasonal components and GARCH modeling. This model is contained in the group of models (see also Cao and Wei, 2000) that rely on a time series approach. The advantage in working with discrete processes is that permit the incorporation of auto-correlation components beyond the lag of one period. Another reason is that the values the temperature indices are calculated are discrete. In such a case it seems better to adopt a discrete process directly, rather than to start with a continuous process and then discretize it. The Campbell-Diebold model specification is given by:

$$T_t = S_t + \sum_{i=1}^{L} r_{t-l} T_{t-l} + u_t \qquad u_t \overset{i.i.d.}{:} N\left(0, s_t^2\right) \tag{3.4a}$$

$$S_t = c_0 + c_1 t + \sum_{i=1}^{P} \left(s_{c,i} \cos\left(2p\, i \frac{d(t)}{365} \right) + s_{s,i} \sin\left(2p\, i \frac{d(t)}{365} \right) \right) \qquad u_t = s_t e_t \tag{3.4b}$$

$$s_t^2 = \sum_{q=1}^{Q} \left(g_{c,q} \cos\left(2p\, q \frac{d(t)}{365} \right) + g_{s,q} \sin\left(2p\, q \frac{d(t)}{365} \right) \right) + a_1 e_{t-1}^2 + b_1 s_{t-1}^2 \qquad e_t \overset{i.i.d}{:} (0,1)^{11} \tag{3.4c}$$

The choice of lags is selected using the Akaike information criteria (AIC) to try to avoid any over-fitting problems that arise from the large number of the estimated coefficients.[12] The extra parameters than need extra explanations are t, a deterministic linear trend, and i, q that control the time frequency of the Fourier series.

The main differences from the BSB Model are that the estimation of the coefficients is done in one step, the significantly higher number of autocorrelation coefficients, the GARCH modeling for the residuals and that the order of the lags and autocorrelation coefficients is based on the AIC. The disadvantage of this model is that is difficult to do analysis with it, as it can be employed for derivatives pricing only by simulation.

3.5. The Principal Components model (PCA)

Principal components analysis (PCA) is a non-parametric technique that

[11] Campbell and Diebold use a modification in the original GARCH dynamics proposed by Engle(1982) and Bollerlev (1986). In our study we employ the original GARCH model as shown in equation (3.4.c)
[12] The maximum order of L, P and Q is 25, 3 and 3 respectively.

summarizes the dynamics of a set of variables by means of a smaller number of variables (principal components-PCs). Stock and Watson (2002) have shown that

PCA can be employed for forecasting purposes. In particular, the PCs are used as predictors in a linear regression equation since they are proven to be consistent estimators of the true latent factors under quite general conditions. Moreover, the forecast constructed from the PCs is shown to converge to the forecast that would be obtained in the case where the latent factors were known. These properties make PCA a very powerful technique for forecasting purposes since it lets the data decide on the predictors to be used.

As the empirical analysis has shown the temperature behaves in different ways in different regions, so we apply PCA to the daily average temperature of our two groups separately. Before applying the technique we detrend and remove the seasonal behavior by regressing the temperature series to trend and Fourier series as in the following equation.[12]

$$T_t = c_0 + c_1 t + \sum_{i=1}^{3}\left(s_{c,i}\cos\left(2p\,i\,\frac{d(t)}{365}\right) + s_{s,i}\sin\left(2p\,i\,\frac{d(t)}{365}\right)\right) + u_t \qquad (3.5a)$$

Next we apply the PCA. We retain the first 6 PCs for the U.S. group and the 4 PCs for the European group. These explain more than 94% of the total variance of the residuals. Next, the fitted residuals are regressed on the previous day values of the first six PCs and first four PCs for the U.S. and European cities, respectively (PCA model). We retain two lags for each component based on the AIC.

$$\hat{u}_{t,us} = r_{0,us} + \sum_{j=1}^{2} r_{1j,us} PC1_{t-j,us} + \sum_{j=1}^{2} r_{2j,us} PC2_{t-j,us} + \sum_{j=1}^{2} r_{3j,us} PC3_{t-j,us} + \sum_{j=1}^{2} r_{4j,us} PC4_{t-j,us} + \sum_{j=1}^{2} r_{5j,us} PC5_{t-j,us} + \sum_{j=1}^{2} r_{6j,us} PC6_{t-j,us} + e_{t,us}$$
$(3.5b)$

$$\hat{u}_{t,eu} = r_{0,eu} + \sum_{j=1}^{2} r_{1j,eu} PC1_{t-j,eu} + \sum_{j=1}^{2} r_{2j,eu} PC2_{t-j,eu} + \sum_{j=1}^{2} r_{3j,eu} PC3_{t-j,eu} + \sum_{j=1}^{2} r_{4j,eu} PC4_{t-j,eu} + e_{t,eu} \qquad (3.5c)$$

Where $r_{i,us}, i = 0,...,6$ and $r_{j,eu}, j = 0,...,4$ are coefficients to be estimated.

3.6. The Equal Weighted Forecast model (EWF)

Equal weighted forecasts are generated by averaging the forecasts of the previous three models, Benth and Saltyte-Benth (BSB), Campbell and Diebold (CD) and principal component analysis (PCA).[13] McNees (1986) argued that the experience of the past tells us that no one forecasting model remains accurate for all variables all the time. The selection of a forecasting model cannot be accomplished by simply choosing the best model for a given period. Given that identifying the best model for each period in most cases is not feasible, combining different forecasts that average differences when one measure gives an over-forecast while the other an under-forecast seems a sensible alternative. Errors may cancel each other out so that the combined forecast would turn out to be relatively closer to the actual value than either forecast independently.

This method does not require any knowledge about the accuracy or the correlation between the errors. Clemen and Winkler (1986) used the simple average of GNP forecasts by utilizing the four major models of Wharton econometrics, the Chase Econometrics, the Data Resources, Inc., and the Bureau of Economic Analysis. Clemen and Winkler (1986) found that the simple average performed better than any single model.

[13] We applied the PCA model to the temperature series directly and we found significant low forecasting ability, sometimes lower than that of the random walk's. Therefore, we do not report results for that model.

4. Out-of-sample forecasting performance

4.1. Evaluation Metrics

Firstly, we briefly review the forecasting methodology, which is rather standard. Specifically, we estimate several models for each series to be forecast and focus on forecast horizons (h) of 1, 5, 10, 15 days. The longer horizon forecasts are produced in a non-overlapping fashion. Our point forecasting proceeds in the rolling scheme.

Let the total sample size be $T+1$. The last P observations of this sample are used for forecast evaluation. The first R observations are used to construct an initial set of regression estimates that are then used for the first prediction. We choose the order of the lags and the order of the Fourier series using the first R observations (***in-sample***). We then forecast the out-of-sample observations by using always a sample of the size R. Namely, the first estimates of the parameters will be estimated with a sample running from 1 to R and then forecast the observation $R+1$, the next estimates with a sample running from 2 to $R+1$ and then forecast the observation $R+2$,…, the last estimates with a sample running from h to $R+h-1$ and then forecast the observation $R+h$.

We assess the out-of-sample forecasting performance of each model that we described in the previous section. The out-of-sample exercise is performed from January 1, 2004 to December 31, 2007.

In line with Campbell and Diebold (2005), we use the root mean squared error (RMSE) to assess the out of sample forecasting performance of the employed models.

i. The root mean squared error (RMSE) is the square root of the average squared deviations of each model's forecast averaged over the number of the forecasts. We use the measure, as our desired model should not produce large forecast errors.

In addition, we use the following metrics:

ii. The mean absolute prediction error (MAE) that is the average of the absolute differences between the forecasts errors of the respective models.

iii. The mean correct prediction (MCP) of the direction of the change that is the average frequency for which the change in temperature predicted by the model has the same sign as the realized change in the temperature.

In order to someone have well-defined results must establish the statistical significance of these measures. So to examine the statistical significance of the RMSE and the MAE we employ the modified Diebold Mariano test of Harvey et al. (1997). In our first evaluation we use as benchmark the random walk model to assess whether any model under consideration outperforms the random walk model. The null hypothesis is that the model under consideration and the benchmark model perform equally well (one-sided test). To compute the Diebold Mariano (DM) test, we use the Newey-West (1987) heteroskedasticity and autocorrelation consistent variance estimator.

Let $d_{m,1} = e_{1,t}^2 - e_{2,t}^2$ and $d_{m,2} = e_{1,t} - e_{2,t}$ where $e_{i,t}, i = 1,2$ is the forecast error of model i.

Given the sequence of P forecasts, Diebold and Mariano (1995) show that

$$P^{1/2}\left(d_{m,1} - m\right) \rightarrow N\left(0, \Omega\right) \text{ where } d_{m,1} = P^{-1}\sum_{t=1}^{P}\left(e_{1,t}^2 - e_{2,t}^2\right).$$ [14]

The test statistic they propose is the following:

$$DM = P^{-1/2} w^{-1/2} \sum_{t=1}^{P}\left(e_{1,t}^2 - e_{2,t}^2\right)$$

Where w is a consistent estimator of the asymptotic variance Ω. Under the null hypothesis, the two non-nested models produce equal RMSEs and DM follows a $N(0,1)$ distribution.

To overcome problems of small out-of-sample forecasts, Harvey et al (1997)

[14] The Diebold-Mariano statistic of the mean absolute error (MAE) is calculated in the same way as that of the root mean squared error (RMSE). We just replace the error loss function of the RMSE with that of MAE.

propose the following modification to the original DM test:

$$MDM = DM * \left\{ \left[P + 1 - 2h + P^{-1} h(h-1) \right] / P \right\}^{1/2}$$

Where modified Diebold-Mariano (MDM) follows the t-distribution with $(P-1)$ degrees of freedom. They also report simulation results revealing that the modified statistic performs better than the original one for forecast horizons, h, greater than 1.

To assess whether any model under consideration outperforms the random walk model in a statistically significant sense under the mean correct prediction (MCP) metric, we use a ratio test. The null hypothesis is that the random walk model and the model under consideration perform equally well (one-sided test).[15]

4.2. One-day Horizon

Tables 3 and 4 report the results for the one day-ahead horizon forecasts of the U.S. and Europe measuring stations, respectively. Specifically, we report the RMSE, the MAE and the MCP for each model and for each weather station for our two groups, respectively. One and two asterisks denote the rejection of the null hypothesis at 1% and 5% significance levels, respectively.

Almost always the employed forecasting models outperform the random walk. Exception to this pattern is the autoregressive model nearly for every city, under the mean absolute error (MAE) metric, and the PCA model, only when employed for the Des Moines and Paris temperature series. Even for the less volatile cities, such as Los Angeles and Barcelona the persistence (RW) forecasts are inadequate, since there exists a statistically predictable pattern in the temperature dynamics (by assuming independence at a level of significance of 1%). The predictable pattern is strongest in the case of the European temperature series than the U.S ones since the accuracy measures are by far smaller.

In Table 4 we observe an interesting result that we should mention. In the northern European cities the temperature models produce more accurate forecasts than

[15] Strictly speaking, the MCP cannot be calculated under the random walk model. Hence, in the ratio test, we treat the random walk model as a naïve model that would yield MCP=50%.

in the southern ones. This result contributes to the empirical analysis inference that the coldest cities are more volatile relative to the hottest ones through the year. We do not deduct the exact same pattern in the U.S., as the weather there is also dependent to extreme phenomena such as the El Nino.[16]

Through the accuracy measures reported in Tables 4 and 5 we observe that the autoregressive model has by far the lowest forecasting ability relative to the other employed models. Given that all models outperform the random walk and the autoregressive model has very low forecasting ability we attempt to answer the question which model performs best among the BSB, CD, PCA and EWF models. We first test pairwise the BSB model using as benchmark models the CD, the PCA and the EWF models sequentially. In the same way we test the CD, PCA and EWF models by using all models sequentially as alternative benchmarks. Overall we calculate twelve modified Diebold-Mariano (one-sided test and 5% significance level) statistics for each city and horizon. In the case where the Null Hypothesis of equal forecasting ability is not rejected in the respective pair of models we report both models. Bold denotes the model with the lowest root mean squared error (RMSE).

Table 5 reports the best model in each city based on the modified Diebold-Mariano (MDM) statistic. We can see that after examining the different forecasting models, no one model was able to outperform the others consistently. The most appropriate forecasting model varied between cities. Thus, a general model that adequately explains the temperature behavior is absent and probably not possible. However, there have been extracted some important inferences.

As far as concerns the U.S. weather stations the PCA model performs best in 60% of the cases. The CD and the EWF models perform best in the rest 40% of the cases. In the case of the European weather stations temperature, the EWF and the PCA models performed best in most cases. Our results reveal that a PCA technique could prove more useful than employing parametric techniques especially in the U.S. weather stations.

[16] El Nino is a set of specific interacting parts of a single global system of coupled ocean-atmosphere climate fluctuations that come about as a consequence of oceanic and atmospheric circulation. The irregularity of E.N. makes predicting a difficult assignment, as it is demonstrably connected to seasonal, even yearly, regional climatic effects on large areas.

4.3. Longer Horizons

We next turn our attention to the accuracy of the forecasting models at longer horizons, namely to 5-, 10- and 15- days ahead. For any given temperature series, we estimate the previously mentioned statistical models with non-overlapping data to check whether there is evidence of predictability in longer horizons.[17] Non-overlapping data are used to avoid the problems in the statistical inference that are encountered in the case of long horizon predictive regressions with overlapping data (see e.g., Valkanov, 2003).

Tables 6 and 7 report the results for the five days-ahead horizon forecasts of the U.S. and Europe measuring stations, respectively. Tables 8 and 9 report the results for the ten days-ahead horizon forecasts of the U.S. and Europe measuring stations, respectively. Tables 10 and 11 report the results for the fifteen days-ahead horizon forecasts of the U.S. and Europe measuring stations, respectively. Specifically, we report the RMSE, the MAE and the MCP for each model and for each weather station for our two groups, respectively. As previously bold denotes rejection of the null hypothesis of equal forecasting ability at the 1% and 5% level.

The forecasts of the employed models outperform almost always those of the random walk's. Compared with the short horizon forecasts their comparative superiority increases with the horizon. As was the case in the one step-ahead horizon forecasts the autoregressive model has the lowest forecasting ability. Table 12 shows the best model per city based on the modified Diebold-Mariano, as in the previous subsection. We can see that there is no single model that yields accurate forecasts for all temperature series, just as was the case with the one day-ahead forecasts. Though, some important deductions are again produced. In the case of the U.S. group as the horizon increases the principal component analysis (PCA) model along with model along with the averaged forecasts (EWF) produce the best forecasts. In the case of the European group as the horizon increases again the EWF and the PCA models perform best. In the longest horizons, fifteen days ahead, the Campbell and Diebold (CD) model performs best almost in every city of our sample. The statistical evidence of a more pronounced predictability in the longest horizons is attributed to the failure

[17] In each horizon, we re-estimate the number of lags for the model of Campbell and Diebold (CD) based on the Akaike information criterion.

of the low order Fourier series to depict the seasonality of the non-overlapping data and to the presence of long memory characteristics..

Quite interestingly, the PCA model we propose leads to better forecasts in many cities and horizons. In addition, as many forecasters advocate the simple averaging of individual models can lead to better forecasts in many cases judging from the performance of the EWF model in the European measuring stations. Finally, our results show also the importance of the incorporation of seasonality in temperature modeling judging from the low performance of the autoregressive model.

To sum up, both at a spatial basis and at a temporal basis, the PCA model in the U.S group and the EWF model in the European group perform best in most cases. If someone wants to forecast temperature in long horizons the model of Campbell and Diebold should be employed. However, the temperature dynamics are not easily explained in a way that many models should be tested in order to choose the best one for a specific area and for a specific time period.

5. Index Forecasting

Weather derivatives are mostly written in terms of the monthly-cumulative sum of the daily HDDs (CumHDDs) and CDDs (CumHDDs), either in a monthly period or in a customer-based period, and traded months prior to the settlement date. Therefore, risk managers and other market participants are required to forecast well in advance estimates of both the CumHDDs and CumCDDs indices to take the appropriate trading positions in the weather market. To address this issue, we attempt to forecast the indices over the same out-of-sample period as the point forecast.

We define heating degree days (HDD_i : measure of cold in winter) and cooling degree days (CDD_i : measure of heat in summer) by the quantities:

$$HDD_i = \max\{65 - T_i, 0\} \quad (5.1)$$

$$CDD_i = \max\{T_i - 65, 0\} \quad (5.2)$$

The cumulative heating degree days (CumHDDs) and the cumulative cooling degree days (CumCDDs) indices for each month are defined by the following equations, where n describes the number of days per month:

$$CumHDD = \sum_{i=1}^{n} HDD_i \text{ and } CumCDD = \sum_{i=1}^{n} CDD_i \quad (5.3)$$

The analysis on the monthly CumHDDs will cover the months of January, February, March, April, October, November and December, while the analysis on the monthly CumCDDs will cover the months of April, May, June, July, August, September and October. The month selection of each weather index is based on the trading specifications of CME.

The forecasted temperatures for the out-of-sample period would be generated using a dynamic forecasting approach of one month. On the last day of each month, and for each city we use the estimated daily model to forecast the temperature of the following month. The forecast for the $t+1$ period is generated using the in sample information included at period t. The forecast for the $t+2$ 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 $t+1$. Then we convert the forecasted temperature into HDD or CDD values, which we cumulate over the HDD or CDD month. After passing through the entire sample, we have 28 observations for the CumHDD and the CumCDD indexes, respectively.

For each of the models - BSB and CD - the forecast accuracy of the forecasted indices is tracked over the same out-of-sample period as the point forecasting one, namely January 2004 to December 31 2007, but now in a monthly accumulated period. [18]

5.2. Index Forecasting Performance

Table 13 reports the descriptive statistics of actual and forecasted values for the out-of-sample periods of the monthly CumHDDs and CumCDDs indices for our two models - BSB and CD. In the first line we report the means for the weather indices and in parenthesis we report the standard deviation of the means. We observe that the standard deviation of the HDD months is higher than that of the CDD months. We do not observe a constant match of the descriptive statistics between the actual values and the forecasted ones of any model. Nevertheless, in some cities we observe an adequate fit between the forecasted values and the actual ones. Figures 3 and 4 plot the realized values for the monthly CumHDDs and CumCDDs indices along with the forecasted values calculated by the best forecasting model in each city. [19] There we observe how well the forecasted series fits the realized one.

Table 14 shows the out-of-sample RMSE and the MAE for our two indices and groups, respectively. In both our groups, significantly higher accuracy measures are associated with the CumHDD index. This implies greater difficulty in forecasting the CumHDD index relative to the CumCDD index. This is probably attributed to the higher weather noise observed during the winter months. Quite interestingly, the

[18] Given the designs of the PCA and the autoregressive models a dynamic step forecast approach is not considered. The EWF model is constructed in this section by averaging the BSB, CD forecasts and the Burn Analysis values.

European weather indices can be forecasted more adequately than those of the U.S. since the accuracy measures are significantly lower. This is attributed to the small variations around the historical mean of the European temperature.

[19] To conserve space we report the figures of two cities from each group.

6. Conclusion

Weather derivatives have gained great popularity in the financial markets as a hedging tool of weather related risk. Markets for weather related products are emerging (CME) and this facts necessities good temperature models to assess the risk exposure. This study has addressed issues whether different temperature models pertain their forecasting power in different areas and in different horizons.

Temperature forecasting is of crucial importance for the market participants of the weather derivatives market. Risk managers want to assess their weather related exposure in order to construct their hedging strategies. Weather derivatives writers need to understand how temperature behaves as standard approaches of arbitrage-free pricing are not applicable in this kind of derivatives. The most rational and practical way to price derivatives is by modeling and forecasting the underlying variable.

In the present study, I have analyzed temperature data measured on ten U.S. and five European stations. After examining five different temperature forecasting models for use as the underlying models for valuing weather derivatives contracts we find that no one model consistently outperforms the others. The temperature behavior is so complex that for each forecast horizon and for each city the best forecasting model varies. However, we tried the principal component analysis (PCA) for forecasting purposes in the case of temperature and we found that it could be a very powerful method. Specifically, in the U.S. cities it beats in most cases its competitors. In the European cities the simple averaging of the individual models (EWF) performed best in most cases. In the longest horizons the Campbell and Diebold model (CD) performed best.

Then we attempted to forecast the cumulative heating degree day (CumHDD) and cumulative cooling degree day (CumCDD) indices so as to see how well the models can forecast the weather indices. This exercise showed that the CumHDD index is more difficult to be forecasted than the CumCDD index. Quite interestingly, the European weather indices can be forecasted more adequately than those of the U.S. since the accuracy measures are significantly lower. This is attributed to the small variations around the historical mean of the European temperature.

In the near future it will be of great interest to attempt to model the weather indices directly. However, that study will require longer temperature series. This may improve the out-of-sample-forecast relative to the daily temperature modeling. Further research should focus also in constructing trading strategies to assess if there is space for speculation purposes using temperature forecasting models.

References

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

Benth F.E., Saltyte-Benth J., 2005. Stochastic Modeling of Temperature Variations with a View towards Weather Derivatives. Applied Mathematical Finance 12 (1), 53-85.

Benth F.E., Saltyte-Benth J., Jalinskas P., 2007. A Spatial-temporal Model for Temperature with Seasonal Variance. Journal of Applied Statistics 34 (7), 823-841.

Benth F.E., Saltyte-Benth J., 2007. The volatility of temperature and pricing of weather derivatives. Quantitative Finance 7 (5), 553-561.

Bollerslev, T., 1986. Generalized ARCH. Journal of Econometrics 31, 307-327.

Bollerslev,T., Wooldridge, J., 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. Econometric Reviews 11, 143-172.

Brockett, P., Wang., M., Yang. C., Zou, H., 2006. Portfolio Effects and Valuation of Weather Derivatives. The Financial Review 41, 55-76.

Brody, D.C., Syroka, J., Zervos, M., 2000. Dynamic pricing of weather derivatives. Quantitative Finance 2, 189-198.

Campbell, S.D., Diebold, F.X., 2005. Weather forecasting for weather derivatives. Journal of the American Statistical Association, 100 (469), 6-16.

Cao, M., Wei, J., 2000. Pricing the weather. Risk 67, 67-70.

Cao, M., Wei, J., 2004. Weather Derivatives Valuation and Market Price of Weather Risk. Journal of Futures Market 24 (11), 1065-1089.

Clemen, R., Winkler, R., 1986. Combining Economic Forecasts. Journal of Business & Economic Statistics 4 (1), 39-46.

Davis, M., 2001. Pricing Weather Derivatives by Marginal Value. Quantitative Finance 1, 305-308.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 256-263.

Dischel, R.S., 2002. Climate Risk and the Weather Market: Financial Risk Management With Weather Hedges, London: Risk Publications.

Dornier, F., Querel, M., 2000. Caution to the Wind. Energy Power Risk Management, 30-32.

Dunis, C.L., Karalis, V., 2003. Weather Derivatives Pricing and Filling Analysis form Missing Temperature Data. Working Paper, Centre for International Banking Economics & Finance.

Engle, R.F., 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987-1008.

Granger, C.W.J., Bates, J.M., 1969. The Combination of Forecasts. Operational Research Quarterly 20 (4), 451-468.

Granger, C.W.J., Ramanathan, R., 1984. Improved Methods of Combining Forecasts. Journal of Forecasting 3 (2), 197-204.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281-291.

Huang, H., Shiu, Y., Lin, P., 2008. HDD and CDD Option Pricing with Market Price of Weather Risk for Taiwan. The Journal of Futures Markets 28 (8), 790-814.

Hull, J., 1999. Options, Futures and Other Derivatives. Fifth Edition, Prentice Hall.

Jewson, S., Brix, A., 2005. Weather Derivatives Valuation. Cambridge University Press, New York.

Kosater, P., 2006. Cross-City Hedging with Weather Derivatives using Bivariate DCC GARCH Models. Discussion Papers in Statistics and Econometrics.

McNees. S., 1986. Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts. Journal of Business & Economic Statistics 4 (1), 5-15.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167-1179.

Stock, J.H., Watson, M.W., 2004. Combining Forecast of Output Growth in a Seven-Country Data Set. Journal of Forecasting 23, 405-430.

Stevenson, M., Oetomo, T., 2005. Hot or Cold? A Comparison of Different Approaches to the Pricing of Weather Derivatives. Journal of Emerging Market Finance 4 (2), 101-133.

Taylor, J.W., Buizza, R., 2006. Density forecasting for Weather Derivative Pricing. Journal of International Forecast 22, 29-42.

Torro, H., Meneu, V., Valor E., 2001. Single Factor Stochastic Models With Seasonality Applied to Underlying Weather Derivatives Variables. Manuscript, University of Valencia.

West, J., Platen, E., 2005. A Fair Pricing Approach to Weather Derivatives. Financial Engineering and the Japanese Markets 11 (1), 23-53.

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

Zeng, L. 2000, Pricing Weather Derivatives. The Journal of Risk Finance 1 (3), 72-78.

Appendices

Appendix I

Weather Derivatives

Weather derivatives are a new class of financial instruments. This instrument enables companies to manage the volumetric risk that results from unfavorable weather patterns.

Traditional derivatives have payoffs that depend on the price of a cluster of fundamentals, such as equities, commodities, currencies or even interest rates. This new kind of derivative's underlying 'asset' is a weather measure, like temperature, precipitation, snowfall, hurricane etc.

The most active participants of the weather derivative market are power and energy companies, since they want to manage their temperature-sensitive revenues effectively. The reason for that is the recent deregulation of the energy market. The producers of energy had long observed the strong correlation between the consumption of energy and the fluctuations of weather. So when the deregulation initiated the market became more competitive and the participants decided to hedge the risk with the use of weather derivatives.

The history of weather derivatives is pretty short. The first weather transaction was executed on the September of 1997, when two American energy companies, Enron and Koch, signed their first weather derivative contract in the over the counter (OTC) market. The reason for this transaction was the expected fluctuations of temperature in the following winter.

Another fact that drove the weather market was the El Nino Southern Oscillation phenomenon and its influence on the economy of the United States. The 'El Nino' part is a fluctuation in the surface temperatures of the eastern half of the equatorial Pacific Ocean, while the 'Southern Oscillation' part is a shift in the wind and pressure patterns over the whole of the Pacific. In the winter of 1998 the phenomenon appeared very intensively and took the attention of media in the States. Then many companies afraid of expected negative results because of the mild winter turned to the weather market.

The most important date of the weather market was the June of 1999, when was established the Weather Risk Management Association (WRMA) by the main marketers of the derivatives. Moreover, one month later the Chicago Mercantile Exchange (CME) created the first organized market of weather derivatives. The CME today lists contracts for the following major cities in the United States: Atlanta, Baltimore, Boston, Chicago, Cincinnati, Dallas, Des Moines, Detroit, Houston, Kansas City, Las Vegas, Minneapolis, New York, Philadelphia, Portland, Sacramento, Salt Lake City and Tucson. The CME lists also contracts for nine European cities, six Canadian and two Asian: Amsterdam, Barcelona, Berlin, Essen, London, Madrid, Paris, Rome, Stockholm, Calgary, Edmonton, Montreal, Toronto, Vancouver, Winnipeg, Osaka and Tokyo. The electronic trading system on CME has attracted new participants, because it provides increased liquidity, low transaction costs and drastic decrease of credit risk.

An important asymmetry in the weather market is the slow development of the European and Asian markets in contrast with the American. Analysts state two main reasons for this fact. The first one is that the European and Asian energy industries are not fully deregulated, while the American is. The second one is the high cost of purchasing the weather data to conduct circumstantial analysis.

I.1 Differences between weather derivatives, financial derivatives and insurance contracts

The main difference between weather derivatives and financial derivatives is that in the first kind the underlying asset is not negotiable. So a trader cannot take simultaneously a position both in the underlying asset and in the derivative. A second important difference is that the underlying asset, weather, cannot be manipulated in contrast with traditional derivatives that has been observed some manipulation, mostly from Hedge Funds. Ending, the vast majority of weather indexes can be forecasted with an adequate level of accuracy.

Many are also the factors where weather derivatives differentiate from the standard insurance products. The main difference is the type of coverage that provide these different tools. Insurance's main activity is to provide protection to extreme physical events, such as hurricanes, earthquakes, fires, etc. Instead, weather derivatives are used to protect the customer from the uncertainty that stems from variations of normal weather conditions. Another important difference is that weather derivatives are more standardized and flexible from insurance products reducing the cost of hedging and increasing the liquidity of the market. A financial derivative is revocable, by taking the opposite position in the market, while an insurance contract does not have this property. Finally, weather contracts can be bought for speculation, like any standard derivatives.

The list of actual trading weather contracts is large and dramatically evolving. The structure of weather derivatives is usually structured as swaps, futures and options based on different underlying weather indexes. The type of measure depends on the specifies of contract and can be based on one single weather variable such as temperature, snowfall, rainfall, wind speed, humidity or a combination of these factors. The most common and developed weather contracts are based on temperature indices. The reason for this development is the previously mentioned correlation between the weather and the energy market. Lastly, temperature is a more easily manageable weather variable than hurricane, snowfall or even precipitation.

I.2 Temperature derivatives

Given a weather station, we note by $T_i^{\max}$ and $T_i^{\min}$, respectively, the maximum and the minimum temperatures (in degree Fahrenheit) measured in one day i. We define the average daily temperature at day i as the average of the day's maximum and minimum temperature on a midnight-to-midnight basis for a specific weather station:

$$T_i = \frac{T_i^{\max} + T_i^{\min}}{2} \tag{I.2.1}$$

Weather derivatives are usually written on the cumulative cooling degree days (CumCDD) or the cumulative heating degree days (CumHDD) over a calendar month or a season. A degree day measures how much a day's average temperature deviates from a threshold (mostly 65^o Fahrenheit). The degree day indices origin's come from the energy sector, since they measure the cooling and heating demand per day.

We define heating degree days (HDD_i : measure of cold in winter) and cooling degree days (CDD_i : measure of heat in summer) by the quantities:

$$HDD_i = \max\{T_{ref} - T_i, 0\} \quad \text{(I.2.2)}$$

$$CDD_i = \max\{T_i - T_{ref}, 0\} \quad \text{(I.2.3)}$$

Where T_{ref} is the reference temperature and T_i the mean temperature of a day i.

In the following table we report a description of the contract specification traded on CME. [1]

	U.S. Weather Derivatives Contract Specification
Contract size	$100 x the CME degree day index
Quotation	Degree day index points
Monthly contracts traded	CDD contract: April, May, June, July August, September, October HDD contract: October, November, December, January, February, March, April
Seasonal contracts traded	CDD contracts: May through September HDD contracts: November through March
Last trading day	Futures: The first exchange business day that is at least two calendar days Options on futures: Same date and time as underlying futures after the contract month
Final settlement price	The exchange will settle the contract to the respective CME degree days Index Reported by the Earth Satellite Corporation

[1] In addition to contracts traded on organized markets (e.g. CME) there are also contracts that are traded over the counter (OTC).

The indices described above define how weather variability is encapsulated for the purposes of a weather derivative contract. The contract is then financially settled using the measured value of the index as the input of a pay-off function. We will consider the pay-off for each of these structures from the point of view of the buyer (long position). The seller of the contract (short position) will have exactly the opposite pay-off. The swap, call and put options are by far the most common and their respective pay-off functions are described below.

The pay-off, p, from a long swap contract is given by:

$$p(x) = \begin{cases} -D(K - L_1) & \text{if } x < L_1 \\ D(x - K) & \text{if } L_1 \le x \le L_2 \\ D(L_2 - K) & \text{if } x > L_2 \end{cases} \qquad \text{(I.2.4)}$$

Where x is the index, D is the tick, K is the strike and L_1 and L_2 are the upper and lower limits expressed in units of the index. OTC contracts are usually traded with limits while the CME contracts do not have limits.[2] A long swap contract has the economic function of insuring against high values of the index by paying the buyer a pay-off dependent on the value of the index. The downside for the buyer of a swap is that he has to pay the seller for low values of the index.

The pay-off, p, from a long call contract is given by:

$$p(x) = \begin{cases} 0 & \text{if } x < K \\ D(x - K) & \text{if } K \le x \le L \\ D(L - K) & \text{if } x > L \end{cases} \qquad \text{(I.2.5)}$$

A long call contract has the economic function of insuring against high values of the index. At the start of the contract the buyer pays a premium to the seller. At the end of the contract the seller pays the buyer a pay-off dependent on the value of the index.

[2] The CME options on futures are European, meaning that can only be exercised at the expiration date.

The pay-off, p, from a long put contract is given by:

$$p(x)=\begin{cases} D(K-L) & \text{if } x<L \\ D(K-x) & \text{if } L\le x\le K \\ 0 & \text{if } x>K \end{cases} \quad \text{(I.2.6)}$$

A long put contract has the economic function of insuring against low values of the index. At the start of the contract the buyer pays a premium to the seller. At the end of the contract the seller pays the buyer a pay-off dependent on the value of the index.

Ending our conversation about temperature derivatives we state the following example for the understanding of their use in practice. An oil company should think that if the following winter will be mild, the company will have low sales. However, if the winter is very cold, the company will have significant incomes. This, in turn, creates volumetric exposure inducing the company to hedge its exposure by selling a heating degree day call contract. If the winter is not particularly cold, the company will keep the premium, while if the winter is too cold the company will have enough money to finance the payout of the option. In this way the oil company reduces its exposure to its weather related volumetric risk.

I.3 Valuation of weather derivatives

Traditional derivatives are priced by no-arbitrage arguments, such as Black-Scholes pricing model, which rest on continuous hedging. This premise cannot be implemented in weather derivatives and hence no riskless portfolio can be built. However, at the first years of the birth of weather derivatives market the first attempts involved the Black-Scholes model. But, quickly this approach was rejected for the following reasons:

- The methodology of Black-Scholes for the valuation of options demands the underlying asset be tradable. Weather indices are not.
- It is impossible to create a risk-neutral portfolio, namely taking positions in the derivative and the underlying asset.
- The weather derivatives do not have a high level of liquidity, yet.

So the market quickly turned to techniques based on historical data. The alternative pricing methodologies that can be followed to obtain the price are the following two, “burn analysis” and “weather modeling”. The first is adopted in the insurance industry while the second one tries to model the weather dynamics.

The “burn analysis” is the most naïve methodology and its main requirement is a good and reliable source of weather data. This method answers to the question: “How much we would have paid if we had sold a similar contract with that we want to price every year in the past?” The average of these prices gives us an indication for the price of our option.

However, this procedure does not take into account the pressures from the demand and the premiums of the underwriters. The procedure is composed by the following steps:

1. Collect the historical weather data.
2. Modify the data to the weather indices.
3. Identify corruptions in the data and correct them.
4. For every year of the past, define the resulting trade payoff.
5. Calculate the average of these payoffs.
6. Discount back from settlement date to today.

The main limitation is that we do not incorporate the weather forecasts. We only assume that the following season will resemble the previous seasons of our sample. Moreover, the literature has not defined an adequate range for the sample of the data for our analysis.

The weather based models try to model and forecast the underlying variable, directly. This procedure consists from the following steps:

1. Collect the historical weather data.
2. Identify corruptions in the data and correct them.
3. Choose a statistical model.
4. Simulate possible weather patterns in the future.
5. Calculate the index for every simulated future pattern.
6. Calculate the payoffs of the derivative.
7. Calculate the average of these payoffs.

8. Discount back from to the settlement date.

I.4 Other Weather Derivatives

In this subsection we will concisely mention the rest of the weather derivatives that are offered from the weather market. The most commonly used after the temperature ones are those that depend on the wind and on the precipitation. However, more specialized weather derivatives exist and their indices depend on variables such as snow depth, snowfall, river flow, hurricanes and even on the number of frost days. Yet, since their use concerns special business fields, these derivatives have very low liquidity. The readers could find more information about them in Dischel (2002) and Jewson (2005).

Appendix II

Gap Filling Methodologies

Here we will discuss more thoroughly the methodologies proposed in the literature about filling missing data.

II.1. Naïve approach

In the simplest approach we fill the missing temperature observation with the same day value of that of previous year.

$$T_t = T_{t-1} \quad \text{(II.1)}$$

Where T_t is the missing temperature data at period t and T_{t-1} is the corresponding observation of the previous year.

II.2. Kosater approach

A more reasonable approach where we fill the missing temperature observation by a weighted temporal approach is that of Kosater; where we fill the missing observation with a weighted average day values of previous and next year's observation.

$$T_t = \left(\frac{T_{t-1} + T_{t+1}}{2} \right) \quad \text{(II.2)}$$

Where T_t is the missing temperature data at period t, T_{t-1} the corresponding observation of the previous year and T_{t+1} the corresponding observation of the next year.

II.3. Campbell and Diebold approach

The approach that Campbell and Diebold (2005) used in their study has the following steps.

1. Earth Sat identifies the geographically closest National Climatic Data Center (NCDC) measuring station for each city.

2. Earth Sat calculates for each city the thirty-year daily average difference of the missing variable (T_{max} or T_{min}) between the measuring station and its reference station. In this calculation, each day in the year is taken as distinct; hence the thirty-year average is based on thirty observations.

3. Finally, Earth Sat adds to the reference station measurement the thirty-year average difference.

II.4. Fallback Approach

Data missing for less than twelve consecutive days

If temperature data is unavailable for less than twelve consecutive days during the calculation period, the missing temperature data will be computed for each observation using a period of the fifteen days immediately prior to, and the fifteen days immediately following the relevant missing observations. If there are unavailable temperature data at the measuring or reference station, within the fifteen days prior to and following to the missing gap then an adjustment period of the first fifteen days on which the relevant temperature values are available on either side of each station should be used to calculate the gap.

Data missing for twelve or more consecutive days

If there are missing temperature observations for more than twelve consecutive days, the filling for each observation will be calculated using temperature data from a period fifteen days prior and fifteen days after the relevant missing observations for the three prior years. If there is no data for the index station or the reference station in the immediate three previous years, the adjustment period will be extended until three years of temperature values can be obtained. In the event that there are unavailable temperatures at the reference station or the index station within the periods of the fifteen days prior to or the fifteen days after the missing temperature observations then an adjustment period should be constructed. This should be based on the first fifteen days on which the relevant temperature data are available (maximum 25 days) on either side of each missing data observations.

II.5. Other Approaches

More complicated methods include the expectation maximization (EM) algorithm, the data augmentation (DA) algorithm, state space models and the Kalman filter, the neural networks regression (NNR) models and the principal component analysis.

Appendix III

Fourier Series

Here we will discuss how the cosine function captures the characteristics of the temperature dynamics.

A Fourier series is an expansion of a periodic function f(x) in terms of an infinite sum of sines and cosines. Fourier series makes use of the orthogonality relationships of the sine and cosine functions. As we add a series of cosine and sine terms (with decreasing amplitudes and decreasing periods), the combined signal resembles the original seasonal variable.

The Fourier series of order P is given by the following specification:

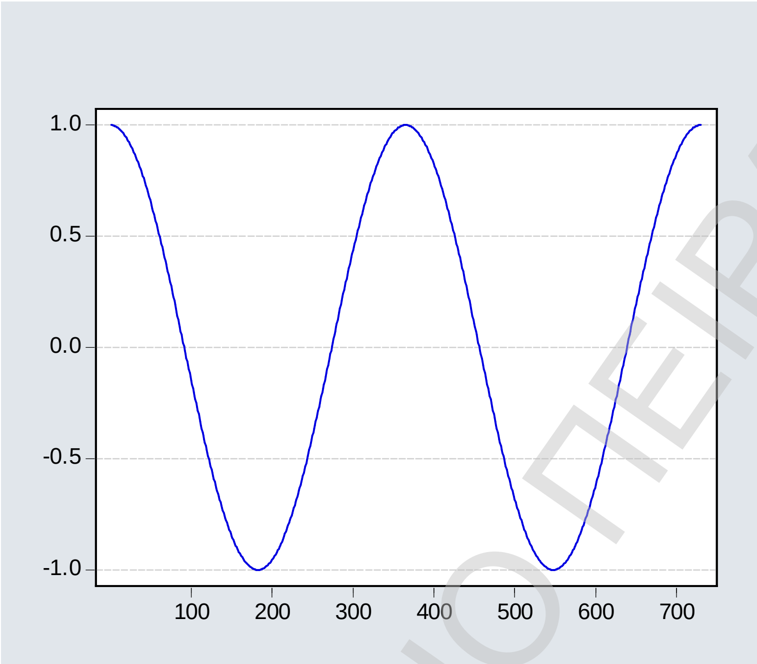
$$\sum_{i=1}^{P}\left(s_{c,p}\cos\left(2p\,p\frac{d(t)}{365}\right)+s_{s,p}\sin\left(2p\,p\frac{d(t)}{365}\right)\right) \qquad \text{(III.1)}$$

Where $s_{c,p}, s_{c,p}$ control the amplitude, p controls the time frequency and $d(t)$ is a step function that cycles through 1, 2,…, 365 (i.e. 1 denotes January 1, 2 denotes January 2 and so on).

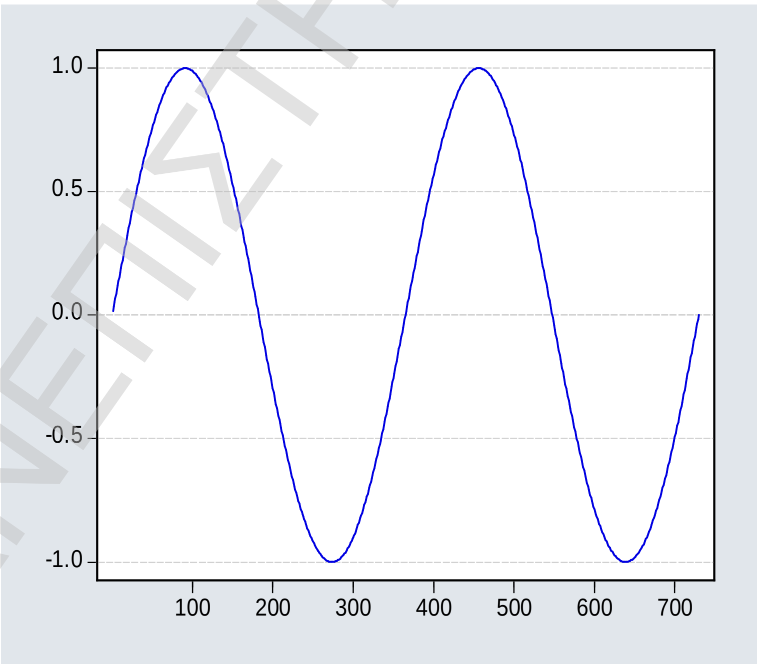
Figure III.1, Panel A and B, shows the movements of the cosine functions for a two-year period, when we replace the time frequency and the amplitude of the functions with the unity. Panel C shows the movements of the previous two functions combined (First order Fourier series).

Figure III.1

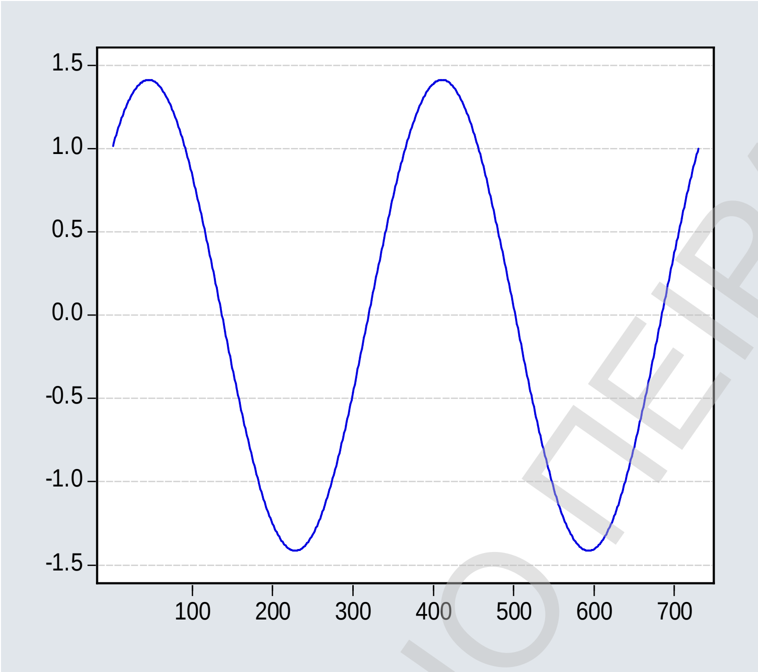
Panel A: $\cos\left(2p\,\frac{d(t)}{365}\right)$



Panel B: $\sin\left(2p\,\frac{d(t)}{365}\right)$



Panel C: $\cos\left(2p\,\frac{d(t)}{365}\right)+\sin\left(2p\,\frac{d(t)}{365}\right)$



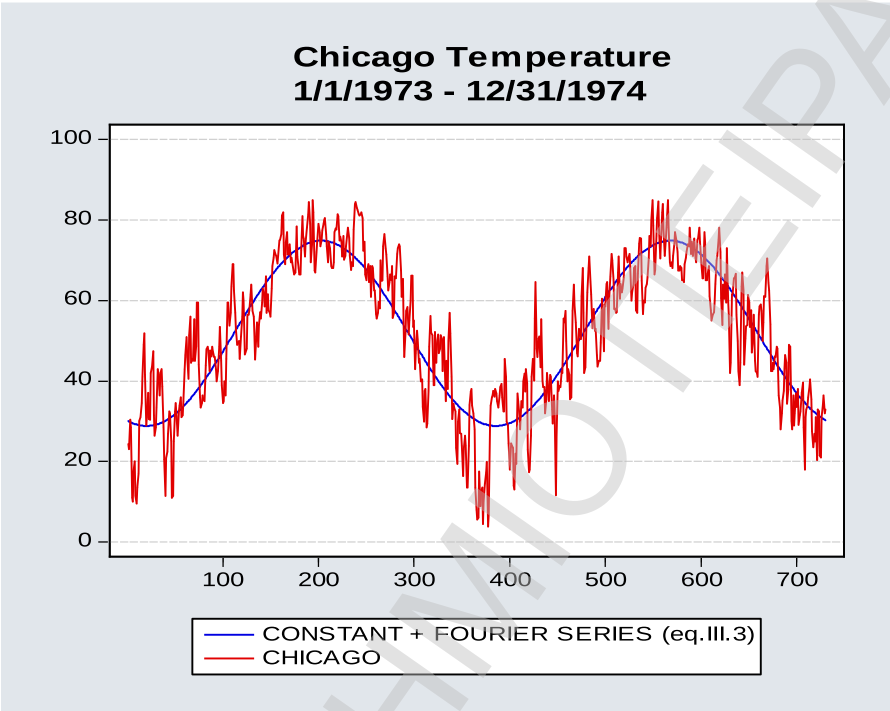
$$T_t = c + s_{c,p}\cos\left(2p\,\frac{d(t)}{365}\right) + s_{s,p}\sin\left(2p\,\frac{d(t)}{365}\right) \qquad \text{(III.2)}$$

To show how well the cosine function captures the seasonality of temperature we regress the temperature series of Chicago (1/1/1973 – 12/31/1974) to a constant, that is interpreted as the average temperature, and the Fourier series of first order as in equation (III.2). After substituting the estimated coefficients of equation (III.2) we reach equation (III.3)

$$T_t = 51.81 - 7.89\cos\left(2p\,\frac{d(t)}{365}\right) - 21.61\sin\left(2p\,\frac{d(t)}{365}\right) \qquad \text{(III.3)}$$

Figure III.2 represents the goodness of the fit of the Fourier series to the temperature of Chicago.

Figure III.2

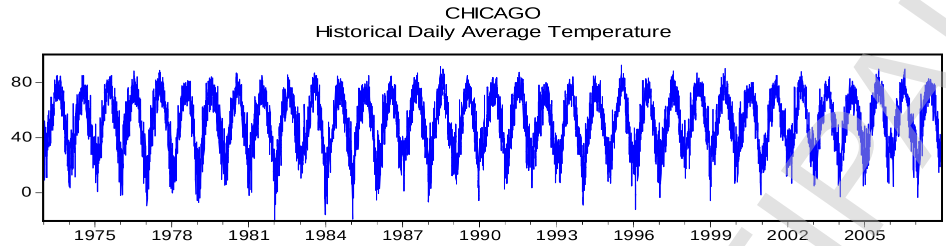


Figures

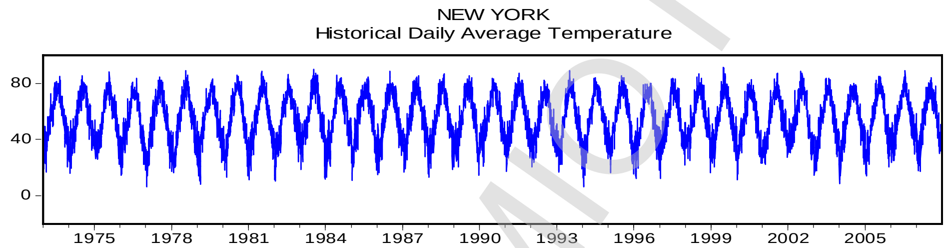
Figure 1: Historical Daily Average Temperature

1-1-1973 to 12-31-2007

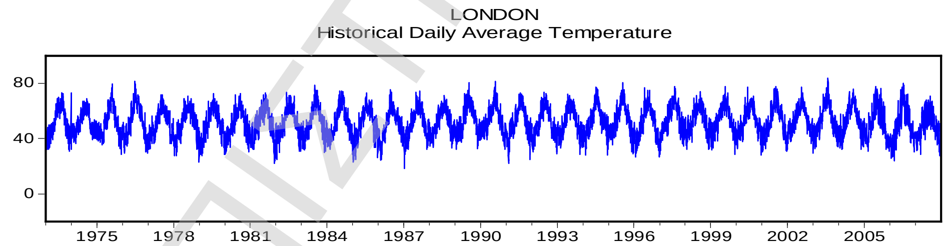
Panel A



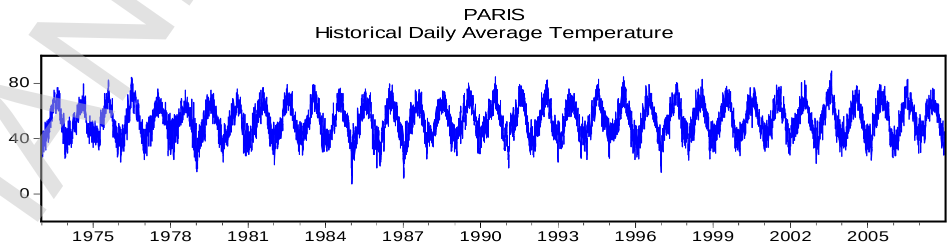
Panel B



Panel C



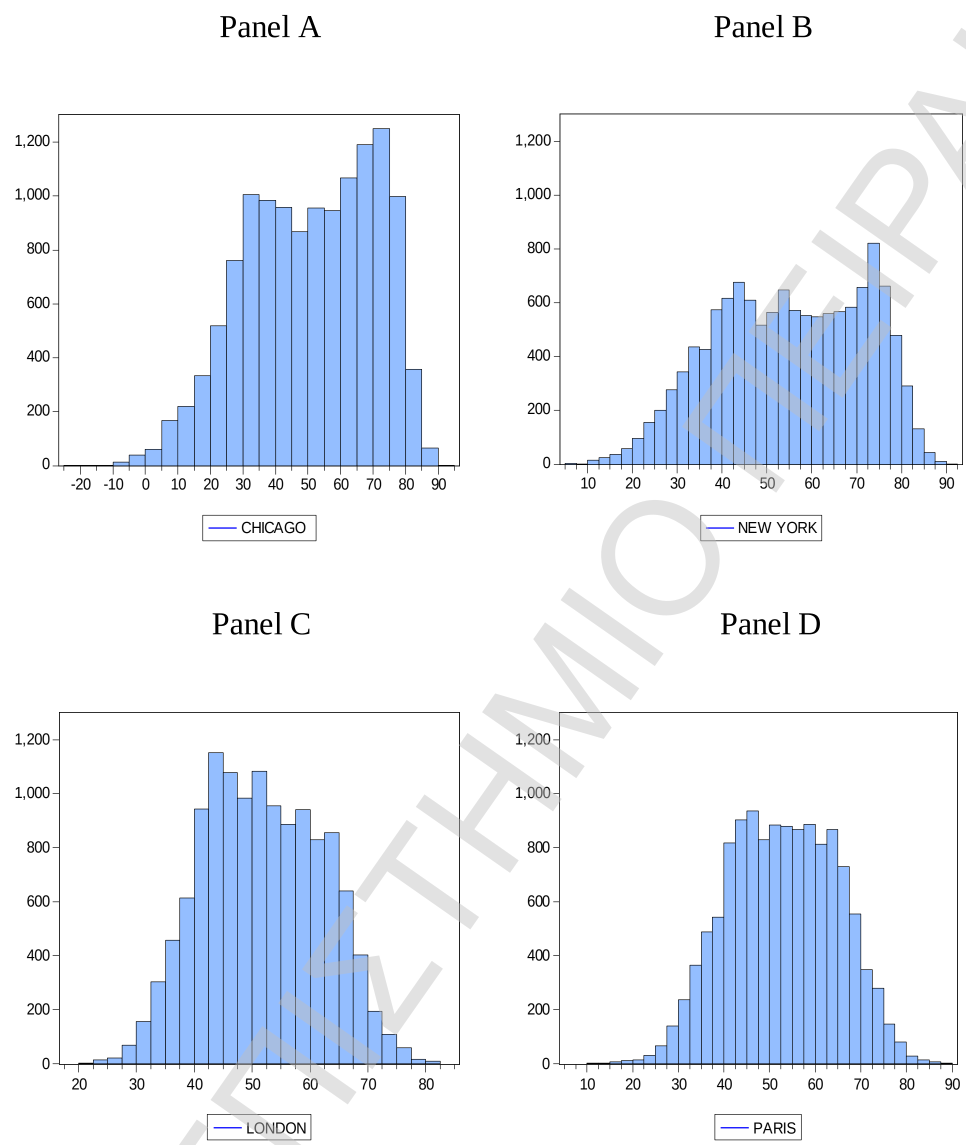
Panel D



Note: Daily Average Temperature is reported in Farenheit Degrees.

Figure 2: Histogram of Daily Average Temperature

1-1-1973 to 12-31-2007

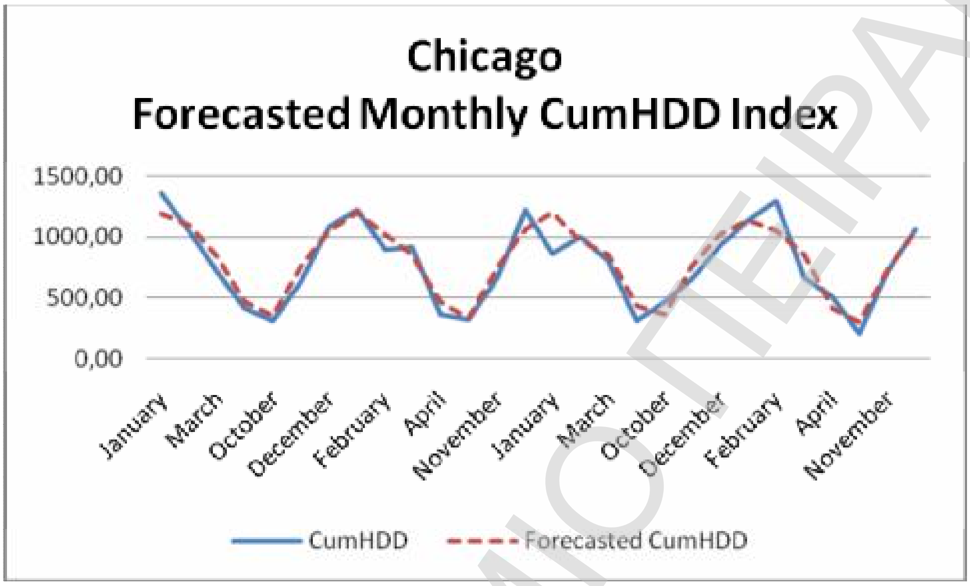


Note: Daily Average Temperature is reported in Farenheit Degrees.

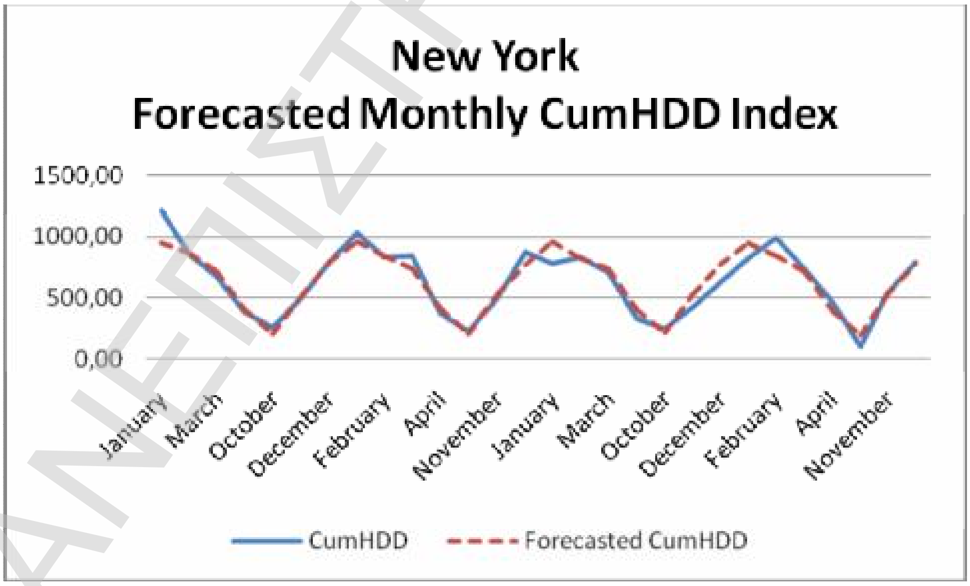
Figure 3: Forecasted Monthly CumHDD Index

January 2004 to December 2007

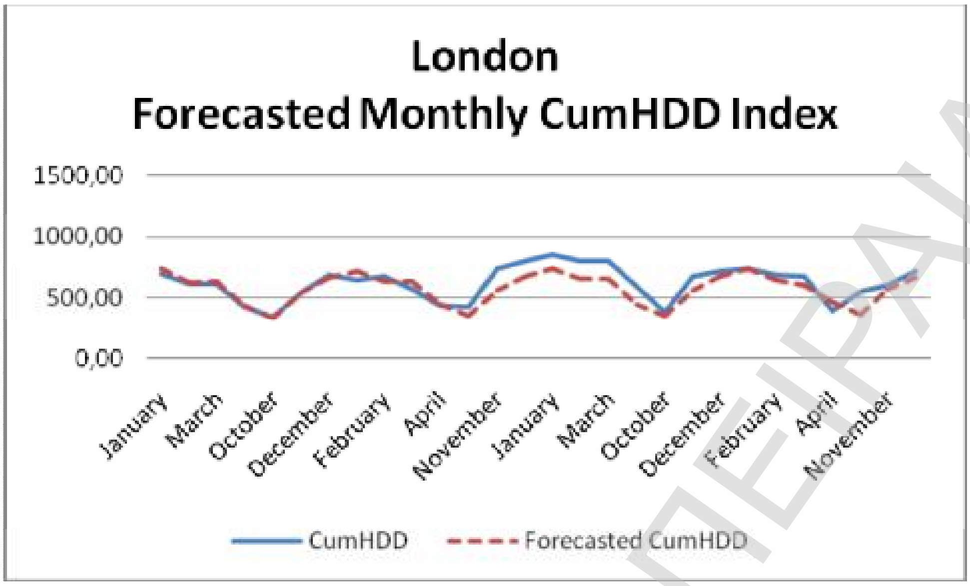
Panel A



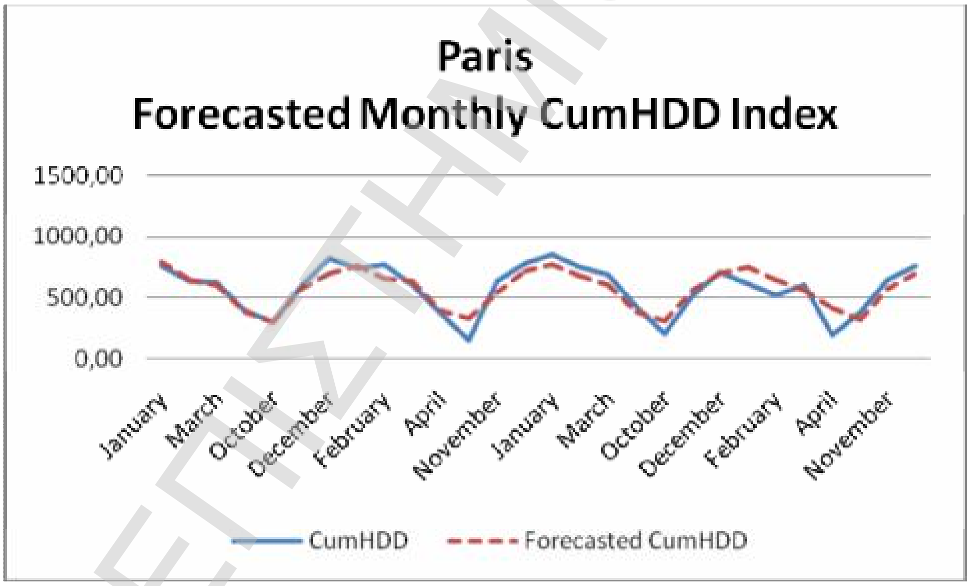
Panel B



Panel C



Panel D

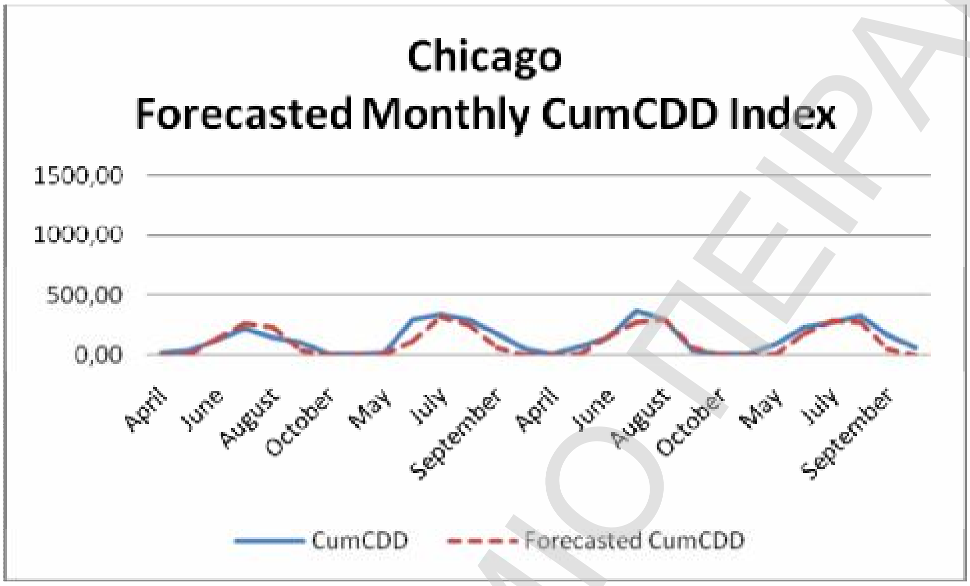


Note: The heating degrees are reported in Farenheit Degrees. The heating degree days (HDD) are calculated with the following equation $HDD_i = \max\{65 - T_i, 0\}$. The monthly cumulative heating degree days are calcualted with the following equation $CumHDD = \sum_{i=1}^{n} HDD_i$, where n denotes the number of days per month. We report in each figure the best forecast, based on the RMSE metric, along with the realized values. The heating months are January, February, March, April, October, November and December.

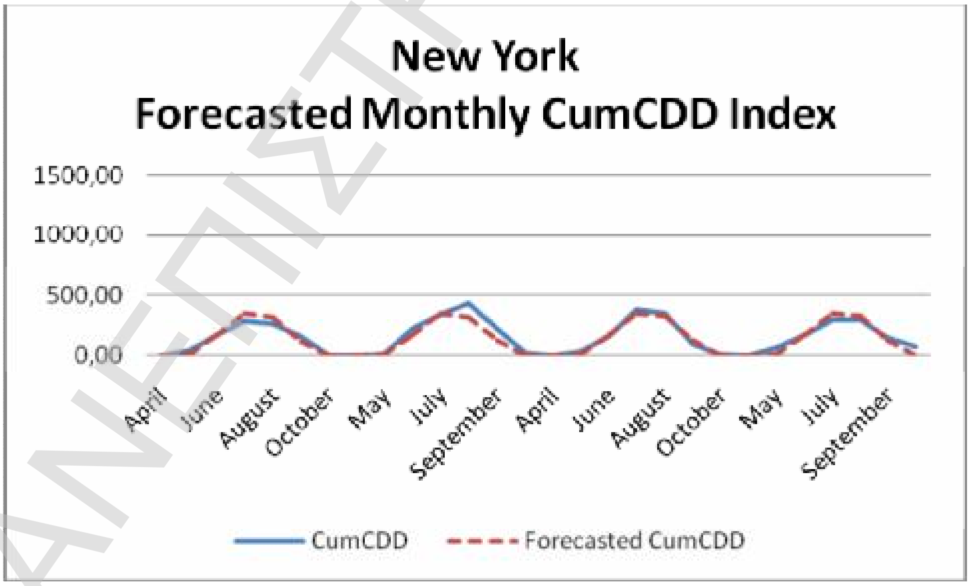
Figure 4: Forecasted Monthly CumCDD Index

April 2004 to October 2007

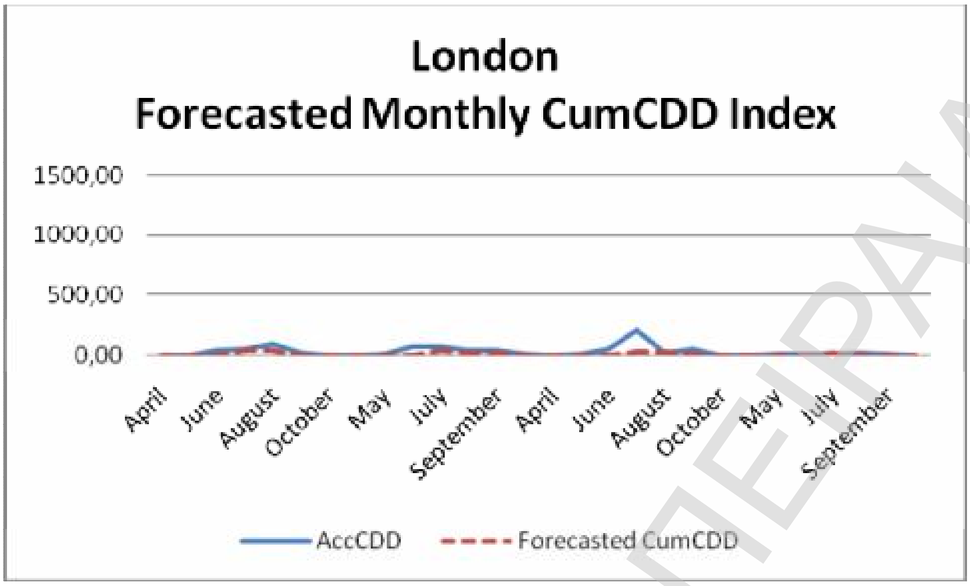
Panel A



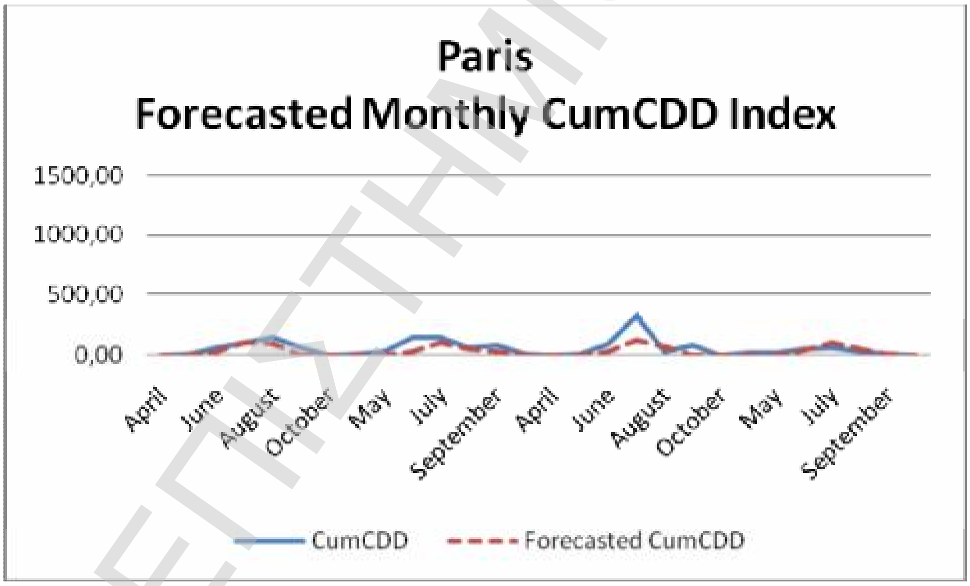
Panel B



Panel C



Panel D



Note: The cooling degrees are reported in Farenheit Degrees. The cooling degree days (CDD) are calculated with the following equation $CDD_i = \max\{T_i - 65, 0\}$. The monthly cumulative cooling degree days are calcualted with the following equation $CumCDD = \sum_{i=1}^{n} CDD_i$, where n denotes the number of days per month. We report in each figure the best forecast, based on the RMSE metric, along with the realized values. The heating months are April, May, June, July, August, September and October.

Tables

Table 1
Temperature Measuring Stations

Panel A: U.S.A.

City	Measuring Station	State
Atlanta	Hartsfield Airport	GA
Chicago	O'Hare Airport	IL
Cincinnati	Convington, KY	OH
Dallas	Dallas - Fort Worth	TX
Des Moines	Des Moines Int.Airport	IA
Las Vegas	McCarran Int.Airport	NV
Los Angeles	Los Angeles Int.Airport	CA
New York	La Guardia	NY
Philadelphia	Philadelphia Int.Airport	PA
Portland	Portland Int.Airport	OR

Panel B: Europe

City	Measuring Station
Barcelona	Barcelona Airport
London	Heathrow Airport
Paris	Paris Airport
Rome	Rome Ciapino
Stockholm	Stockholm Bromma

Table 1: **Temperature Measuring Stations**. The table reports the measuring stations and their respective locations for the U.S. and European cities in panels A and B, respectively. GA denotes the state of Georgia, IL denotes the state of Illinois, OH denotes the state of Ohio, TX denotes the state of Texas, IA denotes the state of Iowa, NV denotes the state of Nevada, CA denotes the state of California, NY denotes the state of New York, PA denotes the state of Pennsylvania and OR denotes the state of Oregon.

Table 2

Summary Statistics for Daily Average Temperature: Jan 1,1973 to Dec 31, 2007

Panel A: U.S.A. (Levels)

	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
Mean	62.68	50.71	56.14	68.00	51.36	68.24	63.45	55.00	56.08	54.67
Median	64.05	52.10	58.00	69.85	53.50	67.55	63.50	55.50	57.05	53.95
Maximum	92.95	92.50	90.50	97.85	91.95	105.45	89.40	91.50	90.95	86.45
Minimum	5.00	-20.05	-10.05	12.10	-17.50	22.45	43.10	5.95	1.95	14.45
Std.Deviation	14.64	19.78	17.66	15.78	20.95	16.86	6.33	16.45	17.37	11.64
Skewness	-0.40	-0.32	-0.37	-0.45	-0.39	0.01	0.05	-0.19	-0.22	0.00
Kurtosis	2.25	2.26	2.27	2.39	2.29	1.77	2.76	2.10	2.06	2.35
ADF(intercept)	**-7.9176***	**-8.2399***	**-8.0467***	**-8.1380***	**-8.2784***	**-8.8637***	**-8.1171***	**-8.7443***	**-8.5929***	**-8.2070***

Panel B: U.S.A. (First Differences)

	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
Mean	-0.0005	0.0002	0.0000	0.0002	0.0000	0.0000	0.0000	-0.0011	-0.0011	-0.0002
Median	0.45	0.40	0.45	0.45	0.45	0.45	0.00	0.00	0.45	0.10
Maximum	25.90	27.45	26.10	23.50	29.95	14.80	16.55	23.50	22.05	15.60
Minimum	-35.45	-32.2	-34.75	-43.30	-28.10	-16.90	-16.45	-28.95	-33.40	-16.45
Std.Deviation	4.57	6.18	6.10	5.75	6.05	3.33	2.46	4.82	5.06	3.43
Skewness	-0.79	-0.20	-0.32	-0.68	-0.22	-0.85	-0.15	-0.15	-0.52	-0.24
Kurtosis	6.31	4.30	4.14	5.43	4.75	5.33	6.12	4.30	4.97	4.09
Jarque-Bera	**7186.98***	**992.55***	**927.27***	**4161.61***	**1749.66***	**4472.94***	**5262.90***	**950.81***	**2653.09***	**768.31***
ADF(intercept)	**-18.2500***	**-16.6009***	**-18.4019***	**-17.4767***	**-16.8205***	**-14.1029***	**-29.7422***	**-14.8178***	**-15.8444***	**-16.8517***

Table 2: Summary Statistics. Entries report the summary statistics of each one of the temperature series in the levels and the first daily differences. The Jarque-Bera and the Augmented Dickey Fuller (ADF) (an intercept has been included in the equation) tests are also reported. One asterisk denotes the rejection of null hypothesis at the 1% level. The null hypothesis for the Jarque-Bera and the ADF Tests is that the series is normally distributed and has a unit root, respectively.

Table 2

Summary Statistics for Daily Average Temperature: Jan 1,1973 to Dec 31, 2007

Panel C: Europe (Levels)					
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
Mean	60.60	51.74	52.75	59.67	43.17
Median	59.35	51.15	52.70	59.00	42.80
Maximum	87.60	84.10	88.80	88.70	80.60
Minimum	29.10	18.50	7.00	22.10	-12.90
Std.Deviation	10.32	10.31	11.91	12.09	14.94
Skewness	0.10	0.06	-0.05	0.05	-0.19
Kurtosis	1.99	2.35	1.96	1.96	2.48
ADF(intercept)	**-8.7624***	**-7.7234***	**-7.7869***	**-8.2529***	**-6.8577***
Panel D: Europe (First Differences)					
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
Mean	-0.0003	0.0009	0.0006	-0.0010	-0.0003
Median	0.00	0.00	0.00	0.00	0.00
Maximum	15.30	24.05	19.55	24.05	35.75
Minimum	-14.40	-33.05	-26.75	-24.95	-37.55
Std.Deviation	2.79	3.89	3.96	3.13	4.45
Skewness	-0.09	0.00	0.02	-0.02	0.00
Kurtosis	4.27	4.67	3.81	5.32	5.40
Jarque-Bera	**881.27***	**1497.52***	**354.68***	**2868.22***	**3070.15***
ADF(intercept)	**-15.4029***	**-30.3883***	**-20.8369***	**-15.7044***	**-34.3988***

Table 2: Summary Statistics. Entries report the summary statistics of each one of the temperature series in the levels and the first daily differences. The Jarque-Bera and the Augmented Dickey Fuller (ADF) (an intercept has been included in the equation) tests are also reported. One asterisk denotes the rejection of null hypothesis at the 1% level. The null hypothesis for the Jarque-Bera and the ADF Tests is that the series is normally distributed and has a unit root, respectively.

Table 3: 1-step-ahead **U.S.A.**

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
Panel A: RW										
RMSE	4.4799	5.7782	5.6635	5.6812	5.7988	3.2699	2.6958	4.6006	5.0471	3.5202
MAE	3.0933	4.2114	4.1013	4.0000	4.2351	2.3510	1.7883	3.3826	3.6109	2.5493
Panel B: AR(1)										
RMSE	4.4243*	5.7154*	5.5890*	5.5833*	5.7416*	3.2553**	2.6357*	4.5570*	4.9947*	3.4807*
MAE	3.1022	4.2495	4.1424	3.9771	4.2702	2.3658	1.7920	3.3830	3.6451	2.5645
MCP	49.38%	49.31%	50.06%	52.94%**	50.75%	48.21%	47.87%	49.72%	48.01%	48.42%
Panel C: BSB										
RMSE	4.2026*	5.4202*	5.3107*	5.2613*	5.4744*	3.1096*	2.5501**	4.2878*	4.7178*	3.2981**
MAE	2.9744*	4.0416*	3.9919*	3.7546*	4.1324**	2.3110**	1.7523	3.2183*	3.4751*	2.4731**
MCP	58.08%*	59.24%*	58.21%*	59.58%*	56.84%*	56.50%*	54.24%*	59.65%*	57.87%*	57.46%*
Panel D: CD										
RMSE	4.0871*	5.2463*	5.1964*	5.1330*	5.3104*	2.9968*	2.4804*	4.1395*	4.5475*	3.1896*
MAE	2.9333**	3.9220*	3.9074*	3.6956*	3.9892*	2.2324*	1.7084*	3.1090*	3.4069*	2.3972*
MCP	59.58%*	61.43%*	60.06%*	61.09%*	59.31%*	60.34%*	55.57%*	61.84%*	59.38%*	60.68%*
Panel E: PCA										
RMSE	3.4750*	4.7141*	4.4019*	5.3992*	5.8533	3.0025*	2.4655*	3.5945*	3.7833*	3.1970*
MAE	2.5706*	3.5764*	3.3923*	4.1262	4.4953	2.2965**	1.7101*	2.7239*	2.8499*	2.4164*
MCP	69.04%*	68.63%*	70.13%*	61.84%*	56.78%*	59.79%*	53.35%*	68.28%*	71.36%*	60.68%*
Panel F: EWF										
RMSE	3.6605*	4.9159*	4.5988*	5.0526*	5.3881*	2.9534*	2.4752*	3.8108*	4.1213*	3.1983*
MAE	2.6068*	3.6612*	3.4541*	3.6736*	4.0719*	2.2058*	1.6981*	2.8397*	3.0548*	2.3969*
MCP	66.23%*	66.30%*	66.57%*	60.82%*	57.87%*	60.20*	54.45%*	65.12%*	67.87%*	60.41%*

Table 3: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1% and 5% significance levels, respectively.

Table 4: 1-step-ahead			**Europe**		
			Panel A: RW		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	2.6593	5.1765	3.7934	2.9613	4.2765
MAE	2.0219	3.6256	2.9882	2.2067	3.2713
			Panel B: AR(1)		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**2.6431****	**5.0516***	**3.7472***	**2.9436****	**4.2474***
MAE	2.0344	**3.5833***	2.9842	2.2258	3.2611
MCP	49.79%	**55.54%***	50.68%	48.90%	48.15%
			Panel C: BSB		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**2.5165***	**4.7765***	**3.6061****	**2.8031***	**4.0499***
MAE	**1.9377***	**3.4522***	**2.8939****	**2.1185***	**3.1613***
MCP	**57.80%***	**59.58%***	**57.60%***	**57.73%***	**54.65%***
			Panel D: CD		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**2.5272***	**4.7973***	**3.5275***	**2.7989***	**4.0180***
MAE	**1.9465***	**3.4780***	**2.8358***	**2.1127***	**3.1199***
MCP	**58.08%***	**60.47%***	**59.38%***	**58.76%***	**55.34%***
			Panel E: PCA		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**2.4249***	**4.8664***	3.8719	**2.6450***	**4.0175***
MAE	**1.8757***	3.6936	3.0491	**1.9952***	**3.1106***
MCP	**61.16%***	**58.08%***	**62.73%***	**61.78%***	**55.82%***
			Panel F: EWF		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**2.4643***	**4.7347***	**3.4213***	**2.7182***	**4.0005***
MAE	**1.8999***	**3.4814***	**2.7438***	**2.0474***	**3.1033***
MCP	**59.65%***	**60.00%***	**62.26%***	**61.57%***	**56.50%***

Table 4: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1% and 5% significance levels, respectively. RW denotes the random walk model, AR(1) the autoregressive model, BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the equal weighted forecast model and PCA denotes the Principal Component Analysis model.

Table 5

Best Model per City

Panel A: U.S.A.		Panel B: Europe	
	1-day		**1-day**
Atlanta	**PCA**	**Barcelona**	**PCA**
Chicago	**PCA**	**London**	**EWF**
Cincinnati	**PCA**	**Paris**	**EWF**
Dallas	**EWF**	**Rome**	**PCA**
Des Moines	**CD**	**Stockholm**	BSB & CD & **EWF** & PCA
Las Vegas	**EWF**		
Los Angeles	CD & EWF & **PCA**		
New York	**PCA**		
Philadelphia	**PCA**		
Portland	**CD** & EWF & PCA		

Table 5: Best Model per City. Entries report the results of the horse race among the employed models, BSB, CD, EWF and PCA. The results are based on the Modified Diebold-Mariano statistic (5% significance level, one-sided test). We first test pairwise the BSB model using as benchmark models the CD, the PCA and the EWF models sequentially. In the same way we test the CD, PCA and EWF models by using all models sequentially as alternative benchmarks. Overall we calculate twelve modified Diebold-Mariano statistics for each city. In the case where the Null Hypothesis of Equal Forecasting Ability is not rejected in the respective pair of models we report both models. BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the Equal Weighted Forecast model and PCA denotes the Principal Component Analysis model. Bold denotes the model with the lowest root mean squared error (RMSE).

Table 6: 5-step-ahead **U.S.A.**

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
Panel A: RW										
RMSE	7.9024	11.3425	10.6452	11.0815	12.7645	6.8942	4.9661	8.4174	9.2514	6.7175
MAE	6.0712	8.9924	8.2650	8.0311	9.8119	5.5046	3.5289	6.6333	7.2542	5.3488
Panel B: AR(1)										
RMSE	**7.6062***	**10.8547***	**10.1405***	**10.3961***	**12.1695***	**6.7671***	**4.5675***	**8.1485***	**8.9259***	**6.4628***
MAE	5.9822	**8.6452***	**8.0748****	**7.6079***	**9.4300***	5.4255	3.4445	**6.4257****	**7.0526****	**5.1427***
MCP	**56.50%****	**59.24%***	52.39%	**58.90%***	**57.19%***	54.45%	52.39%	**56.16%***	54.79%	**55.47%***
Panel C: BSB										
RMSE	**6.0736***	**8.5276***	**8.0949***	**8.0436***	**9.4541***	**5.3102***	**3.8936***	**6.4295***	**7.0879***	**5.1353***
MAE	**4.6559***	**6.6739***	**6.3137***	**6.1387***	**7.3698***	**4.3222***	**2.8101***	**5.1007***	**5.6921***	**4.0196***
MCP	**70.20%***	**78.08%***	**71.57%***	**67.80%***	**70.54%***	**67.12%***	**68.49%***	**72.26%***	**75.00%***	**71.91%***
Panel D: CD										
RMSE	**6.0781***	**8.6008***	**8.1735***	**8.0274***	**9.4098***	**5.3233***	**3.8997***	**6.4515***	**7.1285***	**5.1508***
MAE	**4.6789***	**6.7002***	**6.3708***	**6.0888***	**7.3091***	**4.3195***	**2.8331***	**5.1125***	**5.7421***	**4.0414***
MCP	**69.17%***	**78.42%***	**70.54%***	**66.78%***	**71.23%***	**68.15%***	**67.80%***	**72.60%***	**71.91%***	**71.23%***
Panel E: PCA										
RMSE	**6.0328***	**8.5428***	**8.0255***	**8.0672***	**9.4239***	**5.2209***	**3.8943***	**6.1939***	**6.8467***	**5.1407***
MAE	**4.6084***	**6.7346***	**6.3110***	**6.2090***	**7.3538***	**4.2831***	**2.8180***	**4.8860***	**5.4352***	**4.0437***
MCP	**67.80%***	**78.76%***	**70.89%***	**67.46%***	**71.23%***	**68.83%***	**68.15%***	**72.94%***	**76.71%***	**70.54%***
Panel F: EWF										
RMSE	**6.0116***	**8.5341***	**8.0678***	**8.0236***	**9.4006***	**5.2601***	**3.8874***	**6.3179***	**6.9737***	**5.1391***
MAE	**4.6009***	**6.6886***	**6.2957***	**6.1237***	**7.3173***	**4.2939***	**2.8125***	**5.0090***	**5.5980***	**4.0330***
MCP	**69.17%***	**79.45%***	**71.57%***	**67.12%***	**70.89%***	**68.15%***	**69.17%***	**72.60%***	**74.65%***	**70.89%***

Table 6: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1%

Table 7: 5-step-ahead			Europe		
			Panel A: RW		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	4.7946	7.3849	7.3487	5.2622	8.1923
MAE	3.7356	5.4734	5.8626	4.1239	6.3866
			Panel B: AR(1)		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**4.6875****	**6.9351***	**7.0081***	**5.1594****	**7.0438***
MAE	3.6492	**5.2382***	**5.6313****	4.0453	**5.4994***
MCP	53.76%	**59.93%***	**55.47%****	**56.16%****	**68.83%***
			Panel C: BSB		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**3.8681***	**6.2596***	**5.8071***	**4.2254***	**6.4018***
MAE	**2.9784***	**4.8880***	**4.7675***	**3.3822***	**5.0757***
MCP	**67.12%***	**67.80%***	**72.94%***	**66.43%***	**68.49%***
			Panel D: CD		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**3.8292***	**6.2139***	**5.8117***	**4.2449***	**6.4154***
MAE	**2.9447***	**4.8460***	**4.7595***	**3.3825***	**5.0493***
MCP	**64.38%***	**68.83%***	**71.91%***	**64.38%***	**69.17%***
			Panel E: PCA		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**3.7942***	**6.3843***	**5.7609***	**4.1308***	**6.3900***
MAE	**2.9545***	**5.0104***	**4.7441***	**3.2963***	**5.1027***
MCP	**67.80%***	**68.15%***	**71.23%***	**67.46%***	**68.15%***
			Panel F: EWF		
	Barcelona	**London**	**Paris**	**Rome**	**Stockholm**
RMSE	**3.8108***	**6.2713***	**5.7728***	**4.3858***	**6.3821***
MAE	**2.9407***	**4.9058***	**4.7528***	**3.5612***	**5.0544***
MCP	**65.75%***	**68.83%***	**71.91%***	**71.91%***	**69.17%***

Table 7: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1% and 5% significance levels, respectively. RW denotes the random walk model, AR(1) the autoregressive model, BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the equal weighted forecast model and PCA denotes the Principal Component Analysis model.

Table 8: 10-step-ahead **U.S.A.**

Panel A: RW										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	7.7860	11.4831	10.4129	12.1102	13.8476	7.5694	5.2410	9.0548	10.1253	8.1587
MAE	5.9287	9.2061	8.2376	9.0328	10.4148	5.9736	4.0304	7.2880	8.0702	6.4352
Panel B: AR(1)										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	7.5455	**10.9844***	**9.9404****	**11.2468***	**13.0992***	7.4194	**4.7621***	**8.7201****	**9.6891***	**7.7005***
MAE	5.9200	**8.6411***	**7.9083****	8.7219	**9.8888****	5.7798	**3.7383****	**7.0414****	**7.6988****	**6.1485****
MCP	56.16%	**57.53%****	56.16%	**59.58%****	**60.27%***	56.16%	**59.58%****	54.10%	55.47%	52.05%
Panel C: BSB										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	**5.6216***	**8.3552***	**7.6630***	**7.9606***	**9.5132***	**5.3480***	**3.8653***	**6.2865***	**6.8330***	**5.3897***
MAE	**4.2063***	**6.6502***	**6.1267***	**6.2867***	**7.4570***	**4.4495***	**2.8033***	**4.9838***	**5.4996***	**4.0998***
MCP	**74.65%***	**81.50%***	**77.39%***	**71.23%***	**73.97%***	**71.91%***	**74.65*%**	**78.08*%**	**79.45*%**	**73.97%***
Panel D: CD										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	**5.5796***	**8.3471***	**7.7496***	**7.9136***	**9.4710***	**5.5198***	**3.8428***	**6.4024***	**6.9488***	**5.3980***
MAE	**4.1229***	**6.6488***	**6.1892***	**6.2629***	**7.4537***	**4.5448***	**2.7866***	**5.1060***	**5.6370***	**4.1208***
MCP	**75.34%***	**80.82%***	**76.71%***	**71.91%***	**72.60*%**	**68.49%***	**73.28%***	**77.39*%**	**76.71*%**	**76.02*%**
Panel E: PCA										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	**5.5956***	**8.4082***	**7.5926***	**7.9279***	**9.3913***	**5.4807***	**3.9279***	**6.2803***	**6.7122***	**5.3723***
MAE	**4.2181***	**6.6338***	**6.0504***	**6.3019***	**7.3267***	**4.5239***	**2.8290***	**5.0352***	**5.4885***	**4.1187***
MCP	**75.34%***	**81.50%***	**74.65*%**	**73.97%***	**71.91%***	**67.80%***	**72.60%***	**76.71*%**	**77.39*%**	**72.60%***
Panel F: EWF										
	Atlanta	**Chicago**	**Cincinnati**	**Dallas**	**Des Moines**	**Las Vegas**	**Los Angeles**	**New York**	**Philadelphia**	**Portland**
RMSE	**5.5813***	**8.3542***	**7.6471***	**7.9110***	**9.4344***	**5.4166***	**3.8685***	**6.2962***	**6.8088***	**5.3779***
MAE	**4.1587***	**6.6334***	**6.1038***	**6.2714***	**7.4001***	**4.4965***	**2.7935***	**5.0267***	**5.5180***	**4.1042***
MCP	**76.71%***	**81.50%***	**76.71%***	**70.54%***	**73.28%***	**67.80%***	**73.97%***	**77.39*%**	**78.76*%**	**73.28%***

Table 8: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1%

Table 9: 10-step-ahead Europe

	Barcelona	London	Paris	Rome	Stockholm
Panel A: RW					
RMSE	5.6578	7.9093	8.0825	5.8766	9.4979
MAE	4.4780	6.0982	6.2979	4.7636	7.4719
Panel B: AR(1)					
RMSE	**5.4873****	**7.3353***	**7.6130***	5.7455	**8.9841***
MAE	**4.2853****	**5.5792***	5.9918	4.6077	**7.0725***
MCP	56.16%	**69.86%***	54.79%	52.73%	**60.95%***
Panel C: BSB					
RMSE	**4.2309***	**6.7217***	**5.9346***	**4.4224***	**6.9281***
MAE	**3.3398***	**5.3222***	**4.7815***	**3.5756***	**5.6639***
MCP	**70.54%***	**76.71%***	**69.17%***	**72.60%***	**72.60%***
Panel D: CD					
RMSE	**4.1952***	**6.6944***	**5.9452***	**4.4128***	**7.0143***
MAE	**3.2550***	**5.3119***	**4.7895***	**3.5865***	**5.6906***
MCP	**70.54%***	**76.71%***	**68.49%***	**71.23%***	**71.23%***
Panel E: PCA					
RMSE	**4.2003***	**6.8832***	**5.9151***	**4.3858***	**6.9130***
MAE	**3.3243***	**5.4388***	**4.7713***	**3.5612***	**5.6933***
MCP	**69.17%***	**73.97%***	**67.12%***	**71.91%***	**71.91%***
Panel F: EWF					
RMSE	**4.1912***	**6.7567***	**5.9228***	**4.3945***	**6.9305***
MAE	**3.2814***	**5.3404***	**4.7730***	**3.5676***	**5.6719***
MCP	**70.54%***	**75.34%***	**69.17%***	**72.60%***	**69.86%***

Table 9: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1% and 5% significance levels, respectively. RW denotes the random walk model, AR(1) the autoregressive model, BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the equal weighted forecast model and PCA denotes the Principal Component Analysis model.

Table 10: 15-step-ahead U.S.A.

Panel A: RW

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	8.4536	11.8450	10.6347	10.8508	12.6938	9.0181	5.1903	10.6027	10.7030	8.4494
MAE	6.2556	9.3932	8.2721	8.2000	9.8463	6.8309	3.8701	8.6015	8.6520	6.7000

Panel B: AR(1)

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	8.1327	11.2532**	10.1085	10.1347**	12.0761**	8.7200**	4.6983**	10.0847**	10.2236**	7.9031*
MAE	6.2826	9.0431	8.1552	8.1186	9.6114	6.6596	3.6390	8.1532**	8.3645	6.2342*
MCP	58.76%**	58.76%**	57.73%	55.67%	60.82%**	57.73%	57.73%	57.73%	57.73%	61.85%*

Panel C: BSB

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	8.1032	11.2696**	10.1143	10.0839**	12.0705**	8.6953**	4.6928**	10.0999**	10.2295**	7.8589*
MAE	6.2334	9.0520	8.1476	8.0071	9.5521	6.6597**	3.6615	8.1654**	8.3757	6.2020*
MCP	61.85%*	57.73%	56.70%	54.63%	58.76%**	55.67%	57.73%	58.76%**	57.73%	61.85%*

Panel D: CD

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	6.7109*	8.9841*	8.5423*	8.3709*	10.2715**	5.7512*	3.9765*	7.5274*	7.8896*	5.7956*
MAE	5.2871**	7.0372*	6.9033**	6.4287*	8.1764*	4.6997*	3.0292*	6.1187*	6.4619*	4.6236*
MCP	71.13%*	72.16%*	70.10%*	72.16%*	73.19%*	76.28%*	71.13%*	80.41%*	77.31%*	73.19%*

Panel E: PCA

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	7.1689**	9.2387*	8.9311**	8.6382*	10.4948*	8.2381**	3.9828*	7.9024*	8.4519*	6.6146*
MAE	5.7354	7.3656*	7.2894	6.4998*	8.2895**	6.6939**	3.0359*	6.1459*	6.5739*	5.1852*
MCP	62.88%*	73.19%*	64.94%*	73.19%*	70.10%*	65.97%*	68.04%*	75.25%*	71.13%*	74.22%*

Panel F: EWF

	Atlanta	Chicago	Cincinnati	Dallas	Des Moines	Las Vegas	Los Angeles	New York	Philadelphia	Portland
RMSE	6.6636*	9.1672*	8.4777*	8.4040*	10.1579*	6.9530*	4.0167*	7.9289*	8.1580*	6.3700*
MAE	5.2332*	7.1137*	6.9114*	6.5541*	7.8408*	5.4815*	3.0258*	6.2152*	6.4916*	5.0282*
MCP	67.01%*	73.19%*	69.07%*	73.19%*	73.19%*	72.16%*	71.13%*	78.35%*	74.22%*	77.31%*

Table 8: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1%

Table 11: 15-step-ahead Europe

	Barcelona	London	Paris	Rome	Stockholm
Panel A: RW					
RMSE	5.6577	8.3996	8.9352	6.8259	8.3507
MAE	4.5778	6.5603	7.0798	5.2237	6.8381
Panel B: AR(1)					
RMSE	5.4526	**7.6888***	**8.2795***	6.6018	8.0157
MAE	4.4162	**5.8694***	**6.6220****	5.1622	6.5432
MCP	51.54%	**67.01%***	55.67%	52.57%	51.54%
Panel C: BSB					
RMSE	5.4571	**7.7552***	**8.3420***	6.6273	8.0455
MAE	4.4120	**5.9606***	**6.7003****	5.1282	6.5677
MCP	49.48%	**67.01%***	56.70%	53.60%	52.57%
Panel D: CD					
RMSE	**4.0675***	**6.5893***	**6.2812***	**4.4734***	**6.4478***
MAE	**3.2559***	**5.1863***	**5.2950***	**3.6156***	**5.0819***
MCP	**75.25%***	**76.28%***	**73.19%***	**73.19%***	**69.07%***
Panel E: PCA					
RMSE	**4.9605***	**7.0445***	**7.0935***	6.2253	**7.5142****
MAE	**3.9386***	**5.6302***	**5.6786***	4.8883	6.2041
MCP	**64.94%***	**74.22%***	**68.04%***	**61.85%***	**63.91%***
Panel F: EWF					
RMSE	**4.4493***	6.6283	**6.8059***	**5.4346***	**6.7804***
MAE	**3.5698***	**5.2653***	**5.5055***	**4.2244***	**5.4695***
MCP	**70.10%***	**75.25%***	**73.19%***	**70.10%***	**71.13%***

Table 11: Out-of-sample performance of the model specifications for each one of the average temperature series. The root mean squared prediction error (RMSE), the mean absolute prediction error (MAE) and the mean correct prediction (MCP) of the direction of change in the value of average temperature are reported. The null hypothesis is that the random walk and the model under consideration perform equally well, against the alternative that the model under consideration performs better, have been tested via the Modified Diebold-Mariano test (for RMSE and MAE) and the ratio test (for MCP). One and two asterisks denote rejection of the null hypothesis at 1% and 5% significance levels, respectively. RW denotes the random walk model, AR(1) the autoregressive model, BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the equal weighted forecast model and PCA denotes the Principal Component Analysis model.

Table 12

Best Model per City

	Panel A: U.S.A.		
	5-day	**10-day**	**15-day**
Atlanta	BSB & CD & **EWF** & PCA	BSB & **CD** & EWF & PCA	CD & **EWF** & PCA
Chicago	**BSB** & CD & EWF & PCA	BSB & **CD** & EWF & PCA	**CD** & EWF & PCA
Cincinnati	BSB & CD & EWF & **PCA**	BSB & CD & EWF & **PCA**	**CD** & EWF & PCA
Dallas	BSB & CD & **EWF** & PCA	BSB & CD & **EWF** & PCA	**CD** & EWF & PCA
Des Moines	BSB & CD & **EWF** & PCA	BSB & CD & EWF & **PCA**	**CD** & EWF & PCA
Las Vegas	EWF & **PCA**	**BSB** & EWF	**CD**
Los Angeles	BSB & CD & **EWF** & PCA	BSB & **CD** & EWF & PCA	**CD** & EWF & PCA
New York	**PCA**	BSB & CD & EWF & **PCA**	**CD** & EWF & PCA
Philadelphia	**PCA**	BSB & EWF & **PCA**	**CD** & EWF & PCA
Portland	**BSB** & CD & EWF & PCA	BSB & CD & EWF & **PCA**	**CD**
	Panel B: Europe		
	5-day	**10-day**	**15-day**
Barcelona	BSB & CD & EWF & **PCA**	BSB & CD & **EWF** & PCA	**CD** & EWF
London	BSB & **CD** & EWF	BSB & **CD** & EWF & PCA	**CD** & EWF
Paris	BSB & CD & EWF & **PCA**	BSB & CD & EWF & **PCA**	**CD**
Rome	EWF & **PCA**	BSB & CD & EWF & **PCA**	**CD**
Stockholm	BSB & CD & **EWF** & PCA	BSB & CD & EWF & **PCA**	**CD** & EWF

Table 12: Best Model per City. Entries report the results of the horse race among the employed models, BSB, CD, EWF and PCA. The results are based on the Modified Diebold-Mariano statistic (5% significance level, one-sided test). We first test pairwise the BSB model using as benchmark models the CD, the PCA and the EWF models sequentially. In the same way we test the CD, PCA and EWF models by using all models sequentially as alternative benchmarks. Overall we calculate twelve modified Diebold-Mariano statistics for each city and horizon. In the case where the Null Hypothesis of Equal Forecasting Ability is not rejected in the respective pair of models we report both models. BSB denotes the Benth & Saltyte-Benth model, CD denotes the Campbell & Diebold model, EWF the Equal Weighted Forecast model and PCA denotes the Principal Component Analysis model. Bold denotes the model with the lowest root mean squared error (RMSE).

Table 13

Descriptive Statistics for Out-of-Sample Index Forecasts

Panel A: U.S.A.

	CumHDD			CumCDD		
	Actual	**CD**	**BSB**	**Actual**	**CD**	**BSB**
Atlanta	349.40	338.85	330.69	274.09	262.43	263.86
	(197.35)	*(213.84)*	*(216.86)*	*(171.48)*	*(183.43)*	*(187.51)*
Chicago	779.85	800.33	800.26	138.57	108.52	109.47
	(331.16)	*(316.03)*	*(318.91)*	*(120.94)*	*(124.72)*	*(126.83)*
Cincinnati	614.08	614.66	615.27	173.47	149.90	150.31
	(278.87)	*(279.94)*	*(278.60)*	*(146.48)*	*(146.59)*	*(147.94)*
Dallas	250.03	233.42	230.94	429.32	432.26	428.87
	(189.91)	*(196.21*	*(196.54)*	*(201.31)*	*(221.40)*	*(219.19)*
Des Moines	770.79	759.45	803.66	164.12	139.45	139.48
	(357.63)	*(349.88)*	*(351.38)*	*(138.91)*	*(148.20)*	*(147.93)*
Las Vegas	248.2	248.86	250.52	532.76	510.08	511.74
	(197.25)	*(197.06)*	*(198.87)*	*(288.05)*	*(285.47)*	*(285.85)*
Los Angeles	166.62	170.92	160.96	92.71	72.34	82.09
	(78.83)	*(87.87)*	*(83.59)*	*(76.72)*	*(64.37)*	*(68.89)*
New York	633.77	627.60	627.60	148.76	136.17	135.91
	(276.38)	*(257.18)*	*(254.89)*	*(137.04)*	*(142.57)*	*(142.27)*
Philadelphia	603.02	589.05	584.70	200.75	182.93	184.92
	(283.10)	*(270.12)*	*(266.75)*	*(160.53)*	*(172.70)*	*(173.99)*
Portland	524.71	519.74	521.26	76.21	58.29	59.48
	(178.01)	*(171.83)*	*(169.96)*	*(83.96)*	*(72.86)*	*(74.15)*

Panel B: Europe

	CumHDD			CumCDD		
	Actual	**CD**	**BSB**	**Actual**	**CD**	**BSB**
Barcelona	309.157	284.35	290.24	190.17	173.47	166.52
	(163.05)	*(135.36)*	*(141.27)*	*(143.42)*	*(152.79)*	*(143.46)*
London	622.89	569.78	565.31	28.02	8.56	9.80
	(140.64)	*(122.17)*	*(121.01)*	*(43.59)*	*(14.75)*	*(15.78)*
Paris	571.86	557.65	556.50	54.02	36.81	37.31
	(199.29)	*(153.22)*	*(152.92)*	*(71.35)*	*(51.48)*	*(52.54)*
Rome	408.98	379.52	378.07	177.09	171.74	167.64
	(201.96)	*(170.01)*	*(166.56)*	*(151.56)*	*(161.66)*	*(156.98)*
Stockholm	906.38	925.47	931.66	11.21	0.26	0.33
	(225.22)	*(199.04)*	*(191.85)*	*(22.07)*	*(1.01)*	*(1.26)*

Table 13: Descriptive Statistics. Entries report the descriptive statistics of actual and forecasted values for the out-of-sample period for the monthly-cumulative HDD (CumHDD) and CDD (CumCDD) indices for each model. In the first line we report the means of the aggregated monthly indices and in the second line the standard deviation of the means. Actual denotes the realized values, BSB denotes the Benth & Saltyte-Benth model and CD denotes the Campbell & Diebold model.

Table 14

Accuracy Measures for Out-of-Sample Index Forecasts

Panel A: U.S.A.

		CumHDD		CumCDD	
		CD	**BSB**	**CD**	**BSB**
Atlanta	**RMSE**	70.9695	82.1775	49.5483	64.4140
	MAE	55.8936	69.5223	39.0454	53.6732
Chicago	**RMSE**	108.8824	114.0621	58.4555	64.0806
	MAE	78.2914	82.8599	38.4136	47.0412
Cincinnati	**RMSE**	113.8036	116.6719	52.6175	77.0483
	MAE	81.0580	87.9249	39.1107	56.3042
Dallas	**RMSE**	74.1307	80.1074	64.5210	69.5067
	MAE	46.7473	59.4224	51.01481	58.1816
Des Moines	**RMSE**	122.9246	130.7966	60.1441	64.4633
	MAE	85.8205	98.1309	44.9673	51.8325
Las Vegas	**RMSE**	57.9957	64.1007	58.8996	62.6442
	MAE	46.2630	47.6285	41.1105	54.4406
Los Angeles	**RMSE**	39.8445	42.0901	44.3320	47.1160
	MAE	31.3198	33.7052	27.6602	35.1768
New York	**RMSE**	87.9198	91.9159	37.2771	42.4155
	MAE	60.1414	65.7332	22.2759	29.2544
Philadelphia	**RMSE**	96.2189	101.8163	49.0785	55.5301
	MAE	66.3705	73.4315	33.3475	43.3142
Portland	**RMSE**	67.6037	58.1094	39.7201	31.8682
	MAE	49.4619	47.7945	28.2733	21.4908

Panel B: Europe

		CumHDD		CumCDD	
		CD	**BSB**	**CD**	**BSB**
Barcelona	**RMSE**	57.8036	70.1286	46.0222	45.9830
	MAE	43.1584	57.0803	34.0599	34.1088
London	**RMSE**	81.0145	91.9492	40.3667	38.8639
	MAE	59.0825	72.1307	20.2896	20.2338
Paris	**RMSE**	75.7042	89.9485	54.2982	55.9755
	MAE	56.3229	72.4141	31.5774	35.4848
Rome	**RMSE**	60.9612	82.9348	35.5837	38.2636
	MAE	47.0816	67.0995	26.1102	29.7207
Stockholm	**RMSE**	100.9543	107.8050	22.7953	23.9287
	MAE	74.3320	86.9862	9.3014	10.8851

Table 14: Accuracy measures for out-of-sample index forecasts. The root mean squared prediction error (RMSE) and the mean absolute prediction error (MAE) of the out-of-sample period for the monthly-cumulative HDD (CumHDD) and CDD (CumCDD) indices for each model are reported. BSB denotes the Benth & Saltyte-Benth model and CD denotes the Campbell & Diebold model.