Wang, W., Härdle, W.K., Cabrera, B.L. & Okhrin, O. (2016). Localising temperature risk. Journal of the American Statistical Association, doi: 10.1080/01621459.2016.1180985

Localising temperature risk

Wolfgang Karl Härdle[*], Brenda López Cabrera[†], Ostap Okhrin [‡], Weining Wang[§]

April 15, 2016

Abstract

On the temperature derivative market, modelling temperature volatility is an important issue for pricing and hedging. In order to apply the pricing tools of financial mathematics, one needs to isolate a Gaussian risk factor. A conventional model for temperature dynamics is a stochastic model with seasonality and intertemporal autocorrelation. Empirical work based on seasonality and autocorrelation correction reveals that the obtained residuals are heteroscedastic with a periodic pattern. The object of this research is to estimate this heteroscedastic function so that, after scale normalisation, a pure standardised Gaussian variable appears. Earlier works investigated

[*]Professor at Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz chair of statistics and Director of C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany and School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899. Email:haerdle@wiwi.hu-berlin.de

[†]Assistant professor at the Ladislaus von Bortkiewicz chair of statistics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:lopezcab@wiwi.hu-berlin.de

[‡]Professor of Statistics and Econometrics at the Faculty of Transportation of Dresden University of Technology, 01062 Dresden, Germany. Email:ostap.okhrin@tu-dresden.de

[§]Assistant professor at the Ladislaus von Bortkiewicz chair of statistics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de. King's College London, UK.

temperature risk in different locations and showed that neither parametric component functions nor a local linear smoother with constant smoothing parameter are flexible enough to generally describe the variance process well. Therefore, we consider a local adaptive modelling approach to find, at each time point, an optimal smoothing parameter to locally estimate the seasonality and volatility. Our approach provides a more flexible and accurate fitting procedure for localised temperature risk by achieving nearly normal risk factors. We also employ our model to forecast the temperature in different cities and compare it to a model developed in Campbell and Diebold (2005).

Keywords: Weather derivatives, localising temperature residuals, seasonality, local model selection

JEL classification: G19, G29, G22, N23, N53, Q59

1 Introduction

The pricing of contingent claims based on stochastic dynamics, for example, stocks or FX rates, is well known in financial engineering. An elegant approach to such a pricing task is based on self-financing replication arguments. An essential element of this approach is the tradeability of the underlying. This however does not apply to weather derivatives, contingent on temperature or rain, since the underlying is not tradeable. In this context, the proposed pricing techniques are based on either equilibrium ideas (Horst and Mueller (2007)) or econometric modelling of the underlying dynamics Campbell and Diebold (2005) and Benth, Benth and Koekebakker (2007) followed by risk neutral pricing.

The equilibrium approach relies on assumptions about preferences (with explicitly known functional forms) though. In this study we prefer a phenomenological approach since the underlying (temperature) we consider is of a varying local nature and our analysis aims at understanding the pricing at different locations and different time points around the world.

A time series approach has been taken by Benth et al. (2007), who corrects for seasonality (in mean), then for intertemporal correlation and finally as in Campbell and Diebold (2005), for seasonal variations. After these manipulations, a Gaussian risk factor needs to be isolated in order to apply continuous time pricing techniques, Karatzas and Shreve (2001).

Empirical studies following this econometrical route show evidence that the resulting temperature risk factor deviates severely from Gaussianity, which in turn challenges the pricing tools, Benth, Härdle and López Cabrera (2011). In particular, for Asian cities, like, for example, Kaohsiung (Taiwan), one observes very distinctive non-normality in the form of clearly visible heavy tails caused by extended volatility in peak seasons. This is visible from Figure 1 where a log density plot reveals a non-normal shoulder structure (kurtosis= 3.22, skewness= -0.08, JB= 128.74).

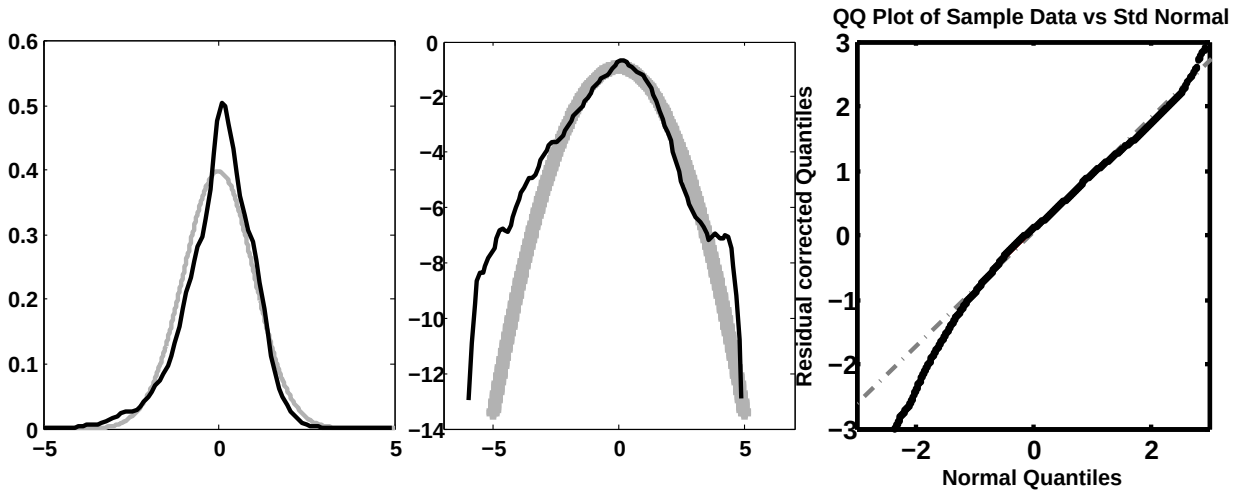


Figure 1: Kernel density estimates (left panel), log kernel density estimates (middle panel) and QQ-plots (right panel) of normal densities (grey lines) and Kaohsiung standardised residuals (black line)

The econometric analysis we apply follows Benth et al. (2007) where temperature is decomposed into a seasonality term and a stochastic part with seasonal variance. The fitted seasonality trend and seasonal variance are approximated with truncated Fourier series (and an additional GARCH term).

The upper panel of Figure 2 displays the seasonality and deseasonalised residuals over two years in Kaohsiung. The lower panel RHS displays the empirical and smoothed seasonal variance function, while the lower panel LHS shows the smoothed seasonal variance function over years. The Fourier series expansion fails though in the volatility peak seasons. Even incorporating an asymmetry term for the dip of temperature in winter does not improve the closeness to normality. One may of course pursue fine tuning the Fourier method with more and more periodic terms but this will increase the number of parameters; we therefore propose a local parametric approach. The mean and the seasonality function estimated with local linear regression using the quartic kernel are also shown in Figure 2. We observe high variance in winter and early summer and low variance in spring and late summer.

The scale correction of the obtained residuals (after seasonal and intertemporal fitting) is apparently not identical over a year. A very structured volatility pattern up to April is followed by a moderately constant period until an increasing peak starting in September. This motivates our research to localise temperature risk. The local smoothness of the seasonal variance function is of course not only a matter of one location (here Kaohsiung) but varies also over the different cities around the world that we are analysing in this study. Our study is local in a double sense: local in time and space. We use adaptive methods to localise the underlying dynamics and with that being able to achieve Gaussian risk factors. This will justify the pricing via standard tools that are based on Gaussian risk drivers. The localisation in time is based on adjusting the smoothing parameter. For a general framework on local parametric approximation we refer to Spokoiny (2009). As a result we obtain better approximations to normality and therefore less biased prices.

This paper is structured as follows. Section 2 describes the localising approach. In section 3, we present the data and conduct the analysis to different cities. Section 4 presents a forecasting exercise and Section 5 is devoted to an application where the pricing of weather derivative contract types is presented. Section 6 concludes the paper. All quotations of currency in this paper will be in USD unless otherwise stated and therefore we will omit the

explicit notion of the currency. All the computations were carried out in Matlab version 7.6 and R. The temperature data for different cities in US, Europe and Asia were obtained from the National Climatic Data Center (NCDC), the Deutscher Wetterdienst (DWD), Bloomberg Professional Service and the Japanese Meteorological Agency (JMA). All data is converted to Celsius degrees. Weather derivative data from CME was extracted from Bloomberg. To simplify notation, dates are denoted using a yyyymmdd format.

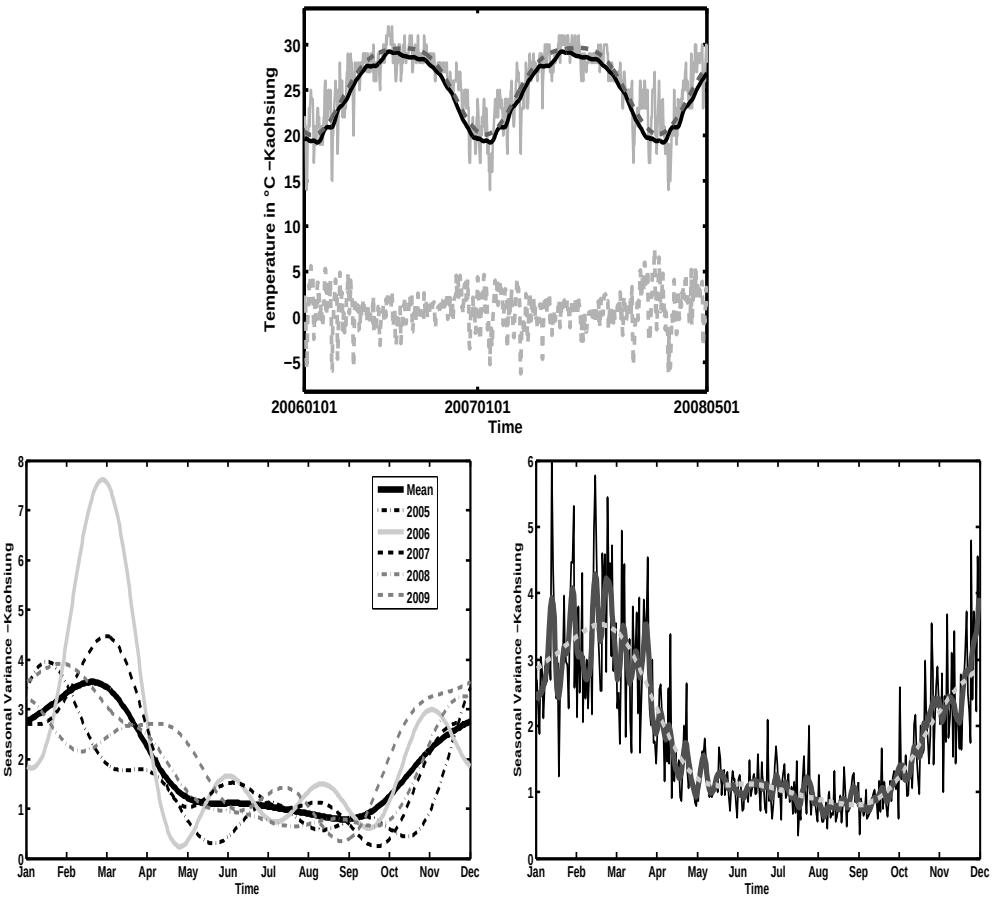


Figure 2: Upper panel: Kaohsiung daily average temperature (grey line), Fourier truncated (dotted grey line) and local linear seasonality function (black line), residuals in lower part. Lower left panel: truncated Fourier seasonal variation ($\hat{\sigma}_t^2$) over years. Lower right panel: Kaohsiung empirical (black line), truncated Fourier (dotted grey line) and local linear (grey line) seasonal variance ($\hat{\sigma}_t^2$) function.

2 Model

Although the temperature data is usually given in a discrete scale, temperature itself develops continuously over time. Thus, a continuous model for the futures price dynamics can be clearly formulated. We propose, as also suggested in Benth et al. (2007) and Härdle and López Cabrera (2012), a mean reverted Ornstein-Uhlenbeck process for the modeling of detrended temperature variations in continuous time $CAR(L)$:

$$d\mathbf{X}_t = \mathbf{A}\mathbf{X}_t dt + \mathbf{e}_L \sigma_t dB_t, \tag{1}$$

where $\sigma_t^2 > 0$ is a bounded deterministic seasonal variation, $\mathbf{X}_t \in \mathbb{R}^L$(detrended temperature) for $L \geq 1$ denotes a vectorial Ornstein-Uhlenbeck process, $\mathbf{e}_k$ a kth unit vector in $\mathbb{R}^L$ for $k = 1, \ldots, L$, B_t a Brownian motion, and an $L \times L$-matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & \ddots & & 0 & \vdots \\ 0 & \ldots & \ldots & 0 & 0 & 1 \\ -\alpha_L & -\alpha_{L-1} & \ldots & & -\alpha_2 & -\alpha_1 \end{pmatrix}.$$

To bring the continuous time model in (1) to data, we consider a discretized version of it. The details of the discretization can be found in the Appendix. Let us first refine our notation from t to (t, j), with $t = 1, \ldots, \tau = 365$ days, $j = 0, \ldots, J$ years. The discrete time series

model for calibration is given as:

$$\begin{aligned} X_{365j+t} &= T_{t,j} - \Lambda_t, \\ X_{365j+t} &= \sum_{l=1}^{L} \beta_{lj} X_{365j+t-l} + \varepsilon_{t,j}, \\ \varepsilon_{t,j} &= \sigma_t e_{t,j}, \\ e_{t,j} &\sim \mathsf{N}(0,1), \\ \hat{\varepsilon}_{t,j} &= X_{365j+t} - \sum_{l=1}^{L} \hat{\beta}_{lj} X_{365j+t-l}, \end{aligned} \tag{2}$$

where $T_{t,j}$ is the temperature at day t in year j, Λ_t denotes the seasonality effect and σ_t the seasonal variance. We adopt the model in (2) and estimate Λ_t, σ_t nonparametricly using adaptive methods proposed later in Section 2.1. Motivation for using this model can be found in Campbell and Diebold (2005) (CD), who proposes the model, see their equations (1), (1a), (1b), (1c).

$$\begin{aligned} T_t &= \text{Trend}_t + \text{Seasonality}_t + \sum_{l=1}^{L} \rho_{t-l} T_{t-l} + \sigma_t \varepsilon_t, \\ \text{Trend}_t &= \sum_{m=0}^{M} \beta_m t^m, \\ \text{Seasonality}_t &= \sum_{p=1}^{P} \left[\delta_{c,p} \cos\left\{2\pi p \frac{d(t)}{365}\right\} + \delta_{s,p} \sin\left\{2\pi p \frac{d(t)}{365}\right\} \right], \\ \sigma_t^2 &= \sum_{q=1}^{Q} \left[\gamma_{c,q} \cos\left\{2\pi q \frac{d(t)}{365}\right\} + \gamma_{s,q} \sin\left\{2\pi q \frac{d(t)}{365}\right\} \right] + \sum_{r=1}^{R} \{\alpha_r (\sigma_{t-r} \varepsilon_{t-r})^2 + \sum_{s=1}^{S} \beta_s \sigma_{t-s}^2\}. \end{aligned}$$

In all the comparisons below, we follow the setting proposed by Campbell and Diebold (2005) with $L = 25, M = 1, P = 3, Q = 3, R = 1$, and $S = 1$. The CD model is also based on a seasonal autoregressive process, but it is quite different from our model in (2). Instead of regressing the deseasonalized temperature on the lagged deseasonalized temperature as in (2), CD model regresses the present's deseasonalized temperature on the temperature in previous days. The trend function thus cannot be interpreted as "seasonal function" but a seasonal

component. Also CD model suggests an additive structure instead of a multiplicative one for the seasonality and GARCH effect in the temperature volatility. Please refer to Benth and Benth (2012) for a detailed discussion of the differences between those two models.

We will use the CD model as a benchmark model for further analysis. Later studies, e.g., Benth et al. (2007) and Härdle and López Cabrera (2012), have provided evidence that the parameters β_{lj} are likely to be j independent and hence estimated consistently from a global autoregressive process model $AR(L_j)$ with $L_j = L$. Also, Benth et al. (2007) adopt the parametrization of Λ_t and σ_t as follows:

$$\Lambda_t = a + bt + \sum_{l=1}^{L_1} c_l \cos\left\{\frac{2\pi(t-d_l)}{l\cdot 365}\right\}, \tag{3}$$

$$\sigma^2_{t,FTSG} = c_{10} + \sum_{l=1}^{L_2}\left\{c_{2l}\cos\left(\frac{2l\pi t}{365}\right) + c_{2l+1}\sin\left(\frac{2l\pi t}{365}\right)\right\} + \alpha_1(\sigma_{t-1}\eta_{t-1})^2 + \beta_1\sigma^2_{t-1}, \tag{4}$$

$$\eta_t \sim iid(0,1).$$

An alternative path to model Λ_t and σ_t is to use a nonparametric method: the local linear regression, where the seasonality Λ_s and σ_s are approximated with a Local Linear Regression (LLR) estimator:

$$\arg\min_{e,f} \sum_{t=1}^{365} \left\{\bar{T}_t - e_s - f_s(t-s)\right\}^2 K\left(\frac{t-s}{h}\right), \tag{5}$$

$$\arg\min_{g,v} \sum_{t=1}^{365} \left\{\hat{\varepsilon}^2_t - g_s - v_s(t-s)\right\}^2 K\left(\frac{t-s}{h}\right), \tag{6}$$

where $\bar{T}_t$ is the mean (over years) of daily averages temperatures, $\hat{\varepsilon}^2_t$ the squared residual process (after seasonal and intertemporal fitting), h the bandwidth and $K(\cdot)$ is a kernel. Note, that due to the spherical character of the data, the kernel weights in (5) and (6) may be calculated from "wrapped around observations" thereby avoiding boundary bias. The estimates $\hat{\Lambda}_s$, $\hat{\sigma}^2_s$ are given by the minimisers $\hat{e}_s$, $\hat{g}_s$ of (5) and (6).

The seasonal trend function Λ_t and the seasonal variance function σ_t^2 affect, of course, the Gaussianity of the resulting normalised residuals. The commonly used approaches 1. truncated Fourier series, and 2. local polynomial regression (with fixed bandwidth) are rather restrictive and do not fit the data well since they do not necessarily yield normal risk factors. These observations motivated us to consider a more flexible approach. The main idea is to fit a local parametric model for the trend and variance with adaptively chosen window sizes. Specifically, we use kernel smoothing and employ an adaptive technique to choose the bandwidth over days. Other examples of this technique can be found in Cízek, Härdle and Spokoiny (2009) and Chen, Härdle and Pigorsch (2010).

It is worth noting that when we bring our model to the data, one can choose to estimate the mean function year by year as $\hat{\Lambda}_{t,j}$ or take the average over years as $\hat{\Lambda}_t$, this is later referred as the separately estimated mean and the jointly estimated mean methods respectively. Regarding the estimate $\hat{\sigma}_t$, an aggregated approach is developed to tackle the problem of losing information when considering estimates at the individual level or averaging mean (variance) functions over time. This approach considers the minimum variance between the aggregation of yearly local function estimates and an optimal local estimate θ^o. Once the sets of local functions have been identified, the aggregated local function can be defined as the weighted average of all the observations in a given time set. Formally, if $\hat{\theta}^j(t)$ is the localised estimation of the variance function σ^2 at time t of year j, the aggregated local function is given by:

$$\hat{\theta}_\omega(t) = \sum_{j=1}^{J} \omega_j \hat{\theta}^j(t). \tag{7}$$

With this aggregation step across J, we give the same weight to all observations, even to observations that were unimportant at the yearly level. Then a reasonable optimised estimate

will be:

$$\arg\min_{\omega} \sum_{j=1}^{J} \sum_{t=1}^{365} \{\hat{\theta}_{\omega}(t) - \hat{\theta}_j^o(t)\}^2 \quad \text{subject to} \quad \Sigma_{j=1}^{J} \omega_j = 1; \omega_j > 0, j = 1, \ldots, J, \tag{8}$$

where the weights are assumed to be exogenous and nonstochastic, and $\hat{\theta}_j^o$ is defined as one of the following: 1 (Locave), $\hat{\theta}_j^o(t) = J^{-1} \sum_{j=1}^{J} \hat{\sigma}_j^2(t)$, the average of seasonal empirical variances over years, 2, (Locsep) $\hat{\theta}_j^o(t) = \hat{\sigma}_j^2(t)$, the yearly empirical variances, 3, one of above two approaches with maximised p-values over a year. One may interpret this normalisation of weights as an optimisation with respect to different frequencies (yearly, daily). In the next subsection we describe the localisation procedures for Λ_t and σ_t, which are going to be elements of estimation methods applied to the temperature data (our summary of the final estimation methods can be found in Table 3).

2.1 Adaptive estimation

In this subsection we introduce adaptive procedures adopted for flexible estimation of Λ_t and σ_t. The time series $T_{t,j}$ are approximated at a fixed time point $s \in [1, 365]$. Our goal is to find a local window that possesses certain optimality properties, to be defined below. Specifically, for a specified weight sequence, we conduct a sequential likelihood ratio test (LRT) to choose an appropriate bandwidth. Different procedures of estimating seasonality and volatility are studied. Suppose that the object to be approximated is the seasonal variance $\theta(t) = \{\sigma_t^2\}$ (Λ_t can be estimated similarly). A weighted maximum likelihood approach is given by:

$$\begin{aligned} \tilde{\theta}_k(s) &\stackrel{\text{def}}{=} \arg\max_{\theta} L\{W^k(s), \theta\} \\ &= \arg\min_{\theta} \sum_{t=1}^{365} \sum_{j=0}^{J} \{\log(2\pi\theta)/2 + \hat{\varepsilon}_{t,j}^2/2\theta\} w(s, t, h_k), \end{aligned} \tag{9}$$

with the "localising scheme" $W^k(s) = \{w(s, 1, h_k), w(s, 2, h_k), \ldots, w(s, 365, h_k)\}^{\top}$, where

$w(s,t,h_k) = h_k^{-1}K\{(s-t)/h_k\}$, $k = 1,\ldots,K$, $h_1 < h_2 < h_3 < \ldots < h_K$ a prescribed sequence of bandwidths, and $K(u) = 15/16(1-u^2)^2\mathbf{I}(|u| \leq 1)$ (quartic kernel).

Define confidence sets with critical values (Critical Values) $\mathfrak{z}_k$ to level α:

$$\mathfrak{E}_{\alpha,k} = \{\theta : L(W^k, \tilde{\theta}_k, \theta) \leq \mathfrak{z}_k\}, \tag{10}$$

where the likelihood ratio is defined as

$$L(W^\ell, \tilde{\theta}_k, \theta) \stackrel{\text{def}}{=} L(W^\ell, \tilde{\theta}_k) - L(W^\ell, \theta). \tag{11}$$

Equipped with confidence sets (10), we launch the Local Model Selection (LMS) algorithm:

Step 1. Fix a point $s \in \{1, 2, \ldots, 365\}$.

Step 2. Start with the smallest interval h_1: $\hat{\theta}_1 = \tilde{\theta}_1$

Step 3. For $k \geq 2$, $\tilde{\theta}_k$ is accepted and $\hat{\theta}_k = \tilde{\theta}_k$ if $\tilde{\theta}_{k-1}$ was accepted and $\tilde{\theta}_k \in \mathfrak{E}_{\alpha,l}, \forall \ell = 1,\ldots,k-1$, i.e.

$$L(W^k, \tilde{\theta}_\ell, \tilde{\theta}_k) \leq \mathfrak{z}_\ell, \forall \ell = 1,\ldots,k-1.$$

Otherwise, set $\hat{\theta}_k = \hat{\theta}_{k-1}$, where $\hat{\theta}_k$ is the latest accepted after first k steps.

Step 4. Define $\hat{k}$ as the kth step we stopped, and $\hat{\theta}_\ell = \tilde{\theta}_{\hat{k}}, \ell \geq k$.

The Critical Values $\mathfrak{z}_\ell$ used in the sequential test above are computed based on the following algorithm:.

Step 1. Consider first $\mathfrak{z}_1$ and let $\mathfrak{z}_2 = \mathfrak{z}_3 = \ldots = \mathfrak{z}_{K-1} = \infty$. This leads to the estimates $\hat{\theta}_k(\mathfrak{z}_1)$

and the value $\mathfrak{z}_1$ is selected as the minimal one for which

$$\sup_{\theta^*} \mathsf{E}_{\theta^*} |L\{W^k, \tilde{\theta}_k, \hat{\theta}_k(\mathfrak{z}_1)\}|^r \leq \frac{\alpha \mathfrak{r}_r}{K-1}, k = 2, \ldots, K. \tag{12}$$

Step 2. Suppose $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$ have been fixed, and set $\mathfrak{z}_k = \ldots = \mathfrak{z}_{K-1} = \infty$. With estimate $\hat{\theta}_m(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)$ for $m = k+1, \ldots, K$. select $\mathfrak{z}_k$ as the minimal value which fulfills

$$\sup_{\theta^*} \mathsf{E}_{\theta^*} |L\{W^m, \tilde{\theta}_m, \hat{\theta}_m(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)\}|^r \leq \frac{k\alpha \mathfrak{r}_r}{K-1} \tag{13}$$

for $m = k+1, \ldots, K$.

Inequality (12) describes the impact of the k Critical Value to the risk, while the factor $\frac{k\alpha}{K-1}$ in (13) ensures that every $\mathfrak{z}_k$ has the same impact. The values of $(\alpha, r, h_1, \ldots, h_K)$ are prespecified hyper-parameters for which robustness and sensitivity issues will be discussed in Section 3.

To be more specific, the explicit solution of (9) is in fact a Nadaraya-Watson estimator:

$$\begin{aligned} \tilde{\theta}_k(s) &= \sum_{t,j} \hat{\varepsilon}_{t,j}^2 w(s,t,h_k) / \sum_{t,j} w(s,t,h_k) \\ &= \sum_{t} \hat{\varepsilon}_{t}^2 w(s,t,h_k) / \sum_{t} w(s,t,h_k), \end{aligned}$$

with

$$\hat{\varepsilon}_t^2 \stackrel{\text{def}}{=} (J+1)^{-1} \sum_{j=0}^{J} \hat{\varepsilon}_{t,j}^2.$$

From a smoothing perspective we are in a comfortable situation here since the boundary bias is not an issue, as we are dealing with a periodic function $\theta(t) = \theta(t + 365)$. We use mirrored observations: assume $h_K < 365/2$, then the observation set, for example for the

seasonal variance, is extended to $\hat{\varepsilon}^2_{-364}, \hat{\varepsilon}^2_{-363}, \ldots, \hat{\varepsilon}^2_0, \hat{\varepsilon}^2_1, \ldots, \hat{\varepsilon}^2_{730}$, where

$$\hat{\varepsilon}^2_t \stackrel{\text{def}}{=} \hat{\varepsilon}^2_{365+t}, -364 \leq t \leq 0,$$
$$\hat{\varepsilon}^2_t \stackrel{\text{def}}{=} \hat{\varepsilon}^2_{t-365}, 366 \leq t \leq 730.$$

Since the location s is fixed, we drop s for simplicity of notation.

The theoretical background for the adaptation procedure can be found in the Appendix.

3 Empirical analysis

We conduct an empirical analysis of temperature patterns for different cities. The main data set contains the daily average temperatures for different cities in Europe, Asia, and the US for the period 1900-2011: Atlanta, Beijing, Berlin, Essen, Houston, Kaohsiung, New York, Osaka, Portland, Taipei, and Tokyo. However as different cities have different data history, for a wider study composed of 1000 cities, a history longer than five years cannot be fulfilled. Moreover, the normality results and forecast performance would be worse for longer histories. We therefore use only up to five years' subsamples. For the sake of brevity, we present, from now on, only the results from four cities: Berlin, Kaohsiung, New York, Tokyo, and detail the other results in the supplementary material. The four cities are from different countries and are quite representative of different types of weather relevant to the interest of weather derivative analysis. Berlin, New York and Tokyo are cities with weather derivatives that are frequently traded, and Kaohsiung is a coasted city with atypical temperature patterns.

We first check seasonality, intertemporal correlation, and seasonal variation. Table 1 provides the coefficients of the Fourier truncated seasonal function (3) for some cities for different time periods. The coefficient a can be seen as the average temperature, the coefficient b as an indicator for a possible trend within a year. The latter coefficients are stable even when the

City	Period	$\hat{a}$	$\hat{b}$	$\hat{c}_1$	$\hat{d}_1$	$\hat{c}_2$	$\hat{d}_2$	$\hat{c}_3$	$\hat{d}_3$
Berlin	(19480101–20080527)	9.2173	0.0000	9.8932	-157.9123	0.2247	261.2850	0.1591	-127.7303
	(19730101–20080527)	9.3050	0.0001	10.0070	-161.2493	0.4601	-66.0530	-0.3723	-416.4776
	(19730101–20080527)	9.3050	0.0001	10.0070	-161.2493	0.4601	-66.0530	-0.3723	-416.4776
	(19830101–20080527)	9.4581	0.0001	10.0969	-161.7129	0.5205	-51.9929	0.3734	42.0874
	(19930101–20080527)	9.5923	0.0002	10.1995	-162.9774	0.6564	-37.1548	0.4241	41.9970
	(20030101–20080527)	9.6948	0.0007	10.1954	-162.3343	0.5554	-43.2293	0.3269	1.5998
Kaohsiung	(19730101–20081231)	24.2289	0.0001	0.9157	-145.6337	-4.0603	-78.1426	-1.0505	10.6041
	(19730101–19821231)	24.4413	0.0001	2.1112	-129.1218	-3.3887	-91.1782	-0.8733	20.0342
	(19830101–19921231)	25.0616	0.0003	2.0181	-135.0527	-2.8400	-89.3952	-1.0128	20.4010
	(19930101–20021231)	25.3227	0.0003	3.9154	-165.7407	-0.7405	-51.4230	-1.1056	19.7340
New York	(19490101–20081204)	53.1473	0.0001	18.6810	-143.4051	-3.3872	271.5072	-0.4203	-16.3125
	(19730101–20081204)	53.6992	0.0001	18.0092	-148.4124	-3.5236	279.6876	-0.4756	-21.8090
	(19730101–19821204)	53.6037	-0.0000	17.7446	-155.2453	-3.7769	289.7932	-0.8326	-4.2257
	(19830101–19921204)	54.8740	-0.0003	17.6924	-152.7461	-3.4245	284.6412	-0.4933	-218.9204
	(19930101–20021204)	53.8050	0.0003	17.6942	-153.3997	-3.4246	285.7958	0.5753	-315.2792
	(20030101–20081204)	52.9177	0.0012	17.8425	-151.2977	-3.8837	287.2022	-0.1290	-216.7298
Tokyo	(19730101–20081231)	15.7415	0.0001	8.9171	-162.3055	-2.5521	-7.8982	-0.7155	-15.0956
	(19730101–19821231)	15.8109	0.0001	9.2855	-162.6268	-1.9157	-16.4305	-0.5907	-13.4789
	(19830101–19921231)	15.4391	0.0004	9.4022	-162.5191	-2.0254	-4.8526	-0.8139	-19.4540
	(19930101–20021231)	16.4284	0.0001	8.8176	-162.2136	-2.1893	-17.7745	-0.7846	-22.2583
	(20030101–20081231)	16.4567	0.0001	8.5504	-162.0298	-2.3157	-18.3324	-0.6843	-16.5381

Table 1: Seasonality estimates $\hat{\Lambda}_t$ of daily average temperature. All coefficients are non-zero at 1% significance level.

estimation is done in a window length of 10 years. In the sense of capturing volatility peak seasons, the right panel of Figure 3 visualises the power of capturing volatility peak seasons by the seasonal local smoother (5) using the quartic kernel over the estimates modeled under Fourier truncated series (3).

After removing the local linear seasonal mean (5) from the daily average temperatures ($X_t = T_t - \Lambda_{t,LLR}$), we check that X_t is a stationary process with the Augmented Dickey-Fuller (ADF) and the KPSS tests. The analysis of the partial autocorrelations and the Akaike Information criterion (AIC) suggest that an $AR(3)$ model fits the temperature evolution well. Table 2 presents the results of the stationarity tests. The temperature data and the smoothed seasonal functions are plotted on the left panel in Figure 3. To show the pattern of the squared residuals after seasonal and intertemporal fitting ($\hat{\varepsilon}^2_{t,j}$), we plot the averaged square residuals over years and show the empirical curves on the right panel in Figure 3. Besides, we have also plotted in Figure 3 the smoothed curves by using the Fourier method

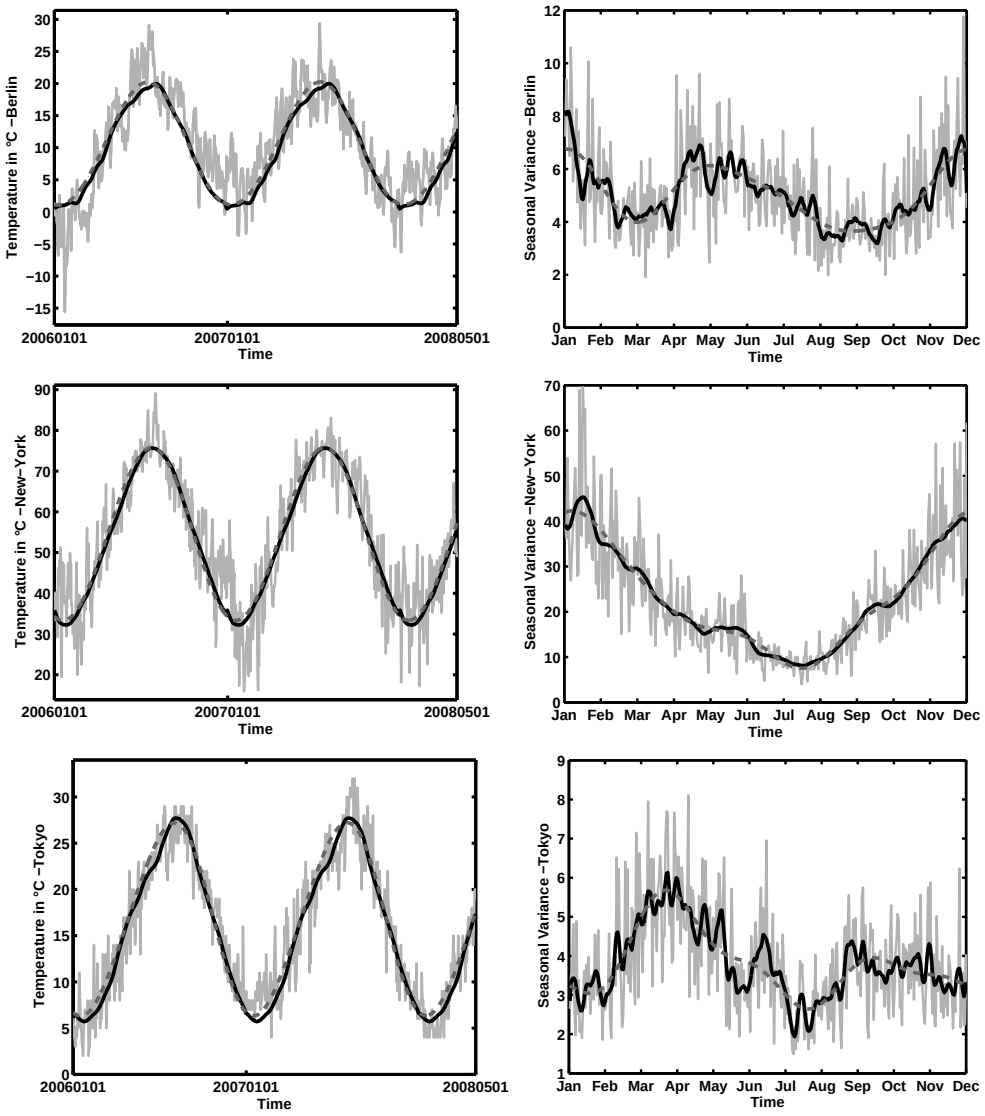


Figure 3: The empirical (grey line), the Fourier truncated (dotted grey line), and the local linear (black line) seasonal mean (left panel) and variance component (right panel) using quartic kernel and bandwidth $h = 4.49$.

City	Period	ADF	KPSS
Atlanta	19480101–20081204	-55.55+	0.21***
Beijing	19730101–20090831	-30.75+	0.16***
Berlin	19480101–20080527	-40.94+	0.13**
Essen	19700101–20090731	-23.87+	0.11*
Houston	19700101–20081204	-38.17+	0.05*
Kaohsiung	19730101–20091210	-37.96+	0.05*
New York	19490101–20081204	-56.88+	0.08*
Osaka	19730101–20090604	-18.65+	0.09*
Portland	19480101–20081204	-45.13+	0.05*
Taipei	19920101–20090806	-32.82+	0.09*
Tokyo	19730101–20090831	-25.93+	0.06*

Table 2: ADF and KPSS-Statistics for the detrended daily average temperature time series for different cities. '+' corresponds to a significance level of 0.01 for ADF test, and '*', '**' and '***' corresponds to significance levels of 0.1, 0.05 and 0.01 respectively for KPSS test.

and the fixed bandwidth local linear method. Furthermore, we check the normality of the final residuals and present the results in the Supplementary Material Tables 1–3 (see there the Fourier method). All seasonal variance estimators lead to residuals that are far from being normally distributed. These facts are of course not an ideal platform for risk neutral pricing (based on standard stochastic financial models). The heavytailedness, as seen in Figure 1, may be attributed to an unsatisfactory extraction of the heteroscedasticity (or mean) function. As a solution we employ a localisation scheme.

The adjustment in the smoothing parameter h will provide the localisation in time. The bandwidth sequences are selected from six candidates: $(1, 2, 3, 4, 5, 6, 7)$, $(1, 2, 3, 5, 7, 10, 13)$, $(3, 5, 7, 9, 11, 13, 15)$, $(3, 5, 8, 12, 17, 23, 30)$, $(5, 7, 10, 14, 19, 25, 32)$, and $(7, 9, 11, 14, 17, 10, 24)$. These candidates are chosen according to the lowest Anderson–Darling (AD) statistic. The best candidate for the bandwidth sequence is the one which yields a residual distribution closest to the normal one. Smoothing the selected bandwidths gives another adaptive estimator, implemented, but not discussed here, due to space limitations.

The Critical Values as calibrated from (12) and (13) are given in Figure 4. The left hand side provides Critical Values simulated from a sample of 10000 observations for a quartic kernel for both mean with $\theta^* = 0$ and variance with $\theta^* = 1$, $r = 0.5$ and different values of

significance level α. The Critical Values for different bandwidth sequences are displayed on the right hand side of Figure 4. The Critical Values, as one observes, are relatively robust to the choice of r and α.

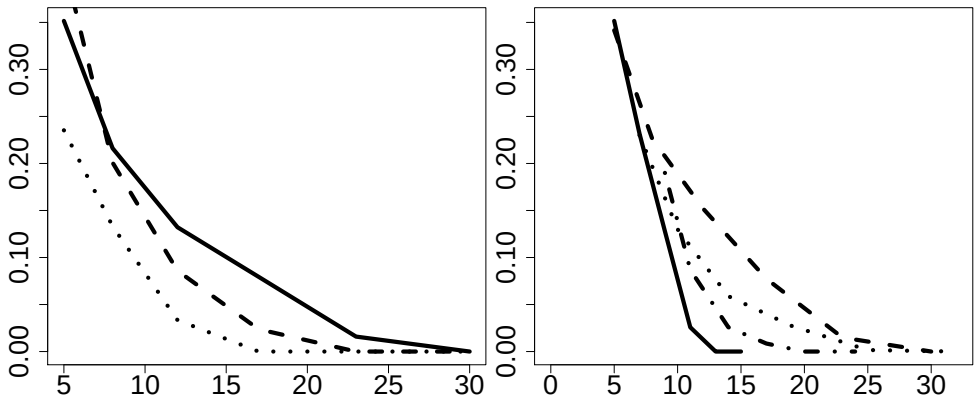
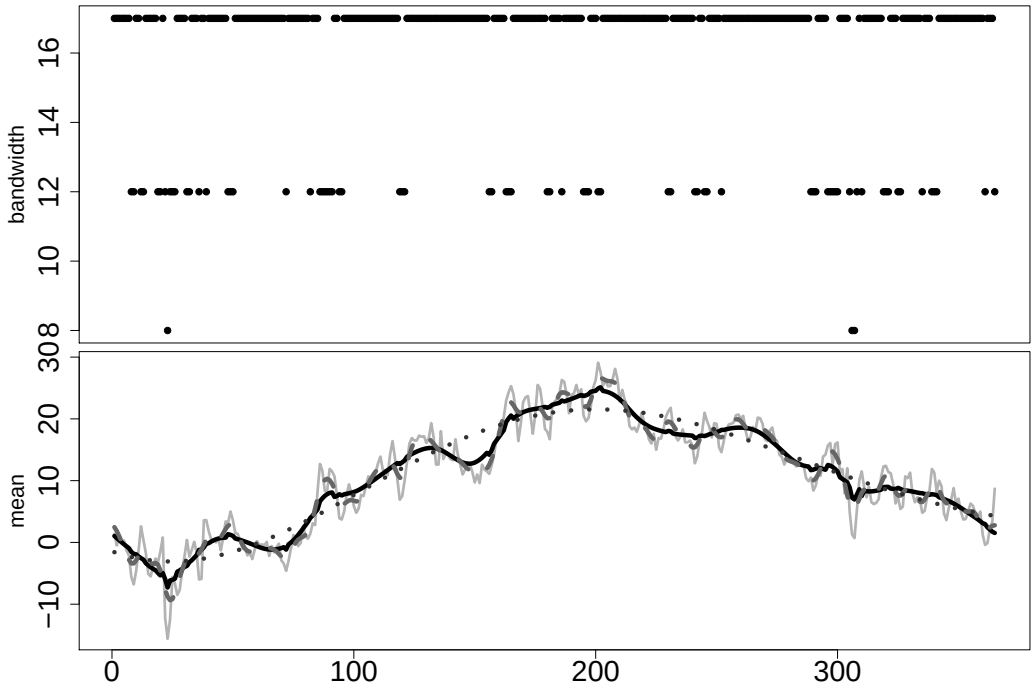


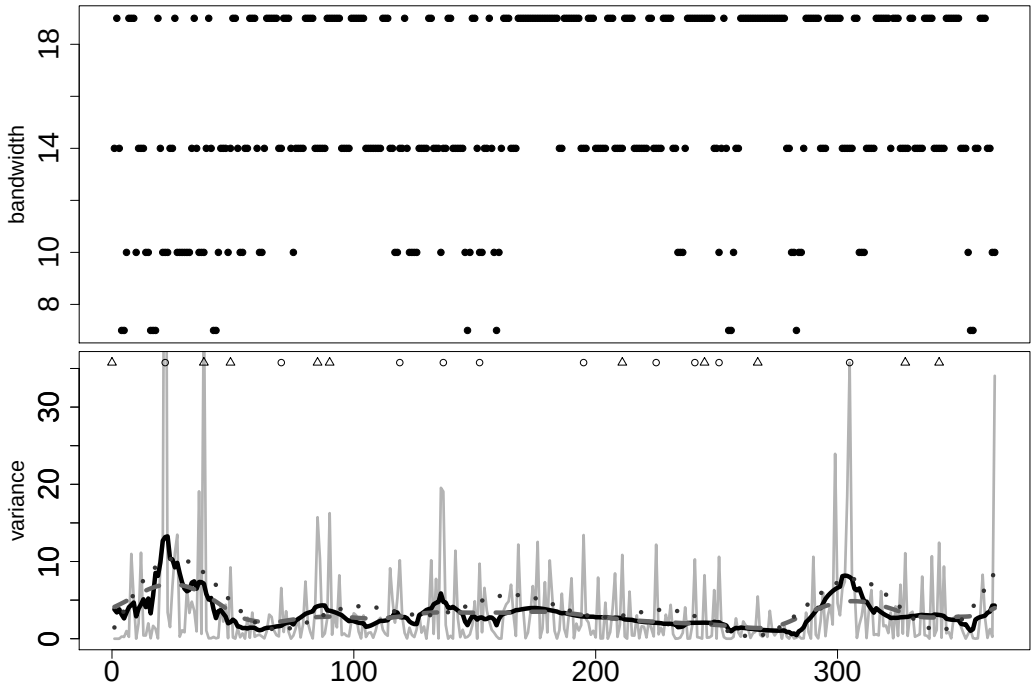
Figure 4: Simulated Critical Values for likelihood of seasonal variance (9) with $\theta^* = 1$, $r = 0.5$, number of simulation runs $= 10000$ with $\alpha = 0.3$ (dotted), 0.5 (dashed), 0.7 (solid) for the bandwidth sequence $(3, 5, 8, 12, 17, 23, 30)$ on the left plot and with $\alpha = 0.3$ and for sequences $(3, 5, 7, 9, 11, 13, 15)$ (solid), $(3, 5, 8, 12, 17, 23, 30)$ (dashed), $(5, 7, 10, 14, 19, 25, 32)$ (dotted), and $(7, 9, 11, 14, 17, 10, 24)$ (dot-dashed) on the right plot.

A one year period is considered in the first place for demonstration purposes, while later we show how the results change with different time periods. Figures 5, 6, 7, and 8 present the general results for the different cities under different adaptive localising schemes for seasonal mean (Me) and seasonal variance (Va): with fixed bandwidth curve (fi), adaptive bandwidth curve (ad), and truncated Fourier (Fourier) for different time intervals. The seasonal mean is estimated jointly over the years, using $\alpha = 0.7$ and power level $r = 0.5$.

The upper panel of each variance plot in Figures 5–8 shows the sequence of bandwidths; the bottom panel displays variance estimation with fixed bandwidth (dashed line), the Fourier truncated method (dotted line), and adaptive bandwidth (solid black line). In all countries, one observes significant differences between the estimates. In particular, in cities like Kaohsiung and New York, one observes more variation of the seasonal variance curves during peak seasons (winter and summer times). The triangles and circles in the bottom panel of each variance plot help us trace the source of the non-normality over time, since they correspond

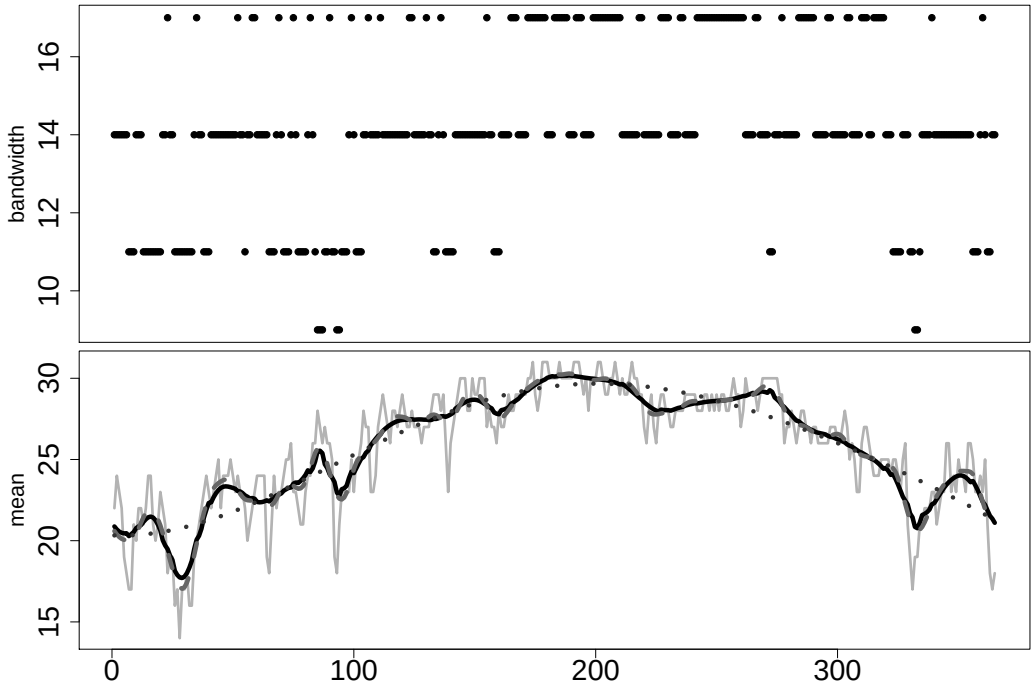


(a) Mean, 2007

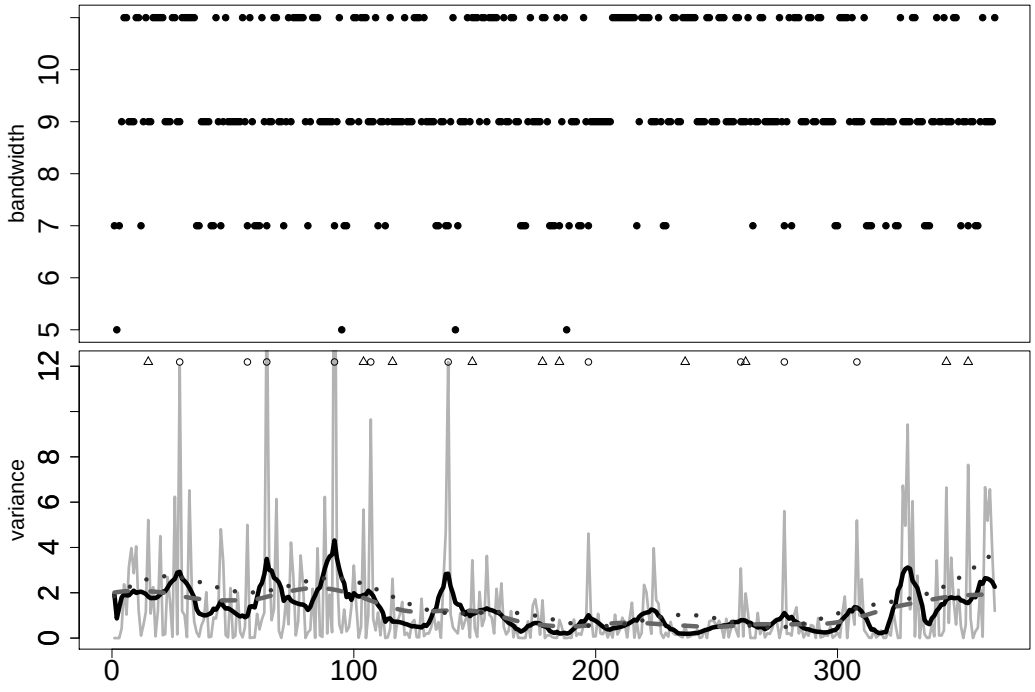


(b) Variance, 2007

Figure 5: Estimation of mean and variance for Berlin. In both figure sequence of bandwidths (upper panel), averaged observations (solid grey line), nonparametric function estimation with fixed bandwidth (dashed grey line), adaptive bandwidth (solid black line) and truncated Fourier (dotted line) (bottom panel of each figure). Circles and triangles in each bottom panel for variance represents the 10 smallest and the 10 largest outliers respectively.

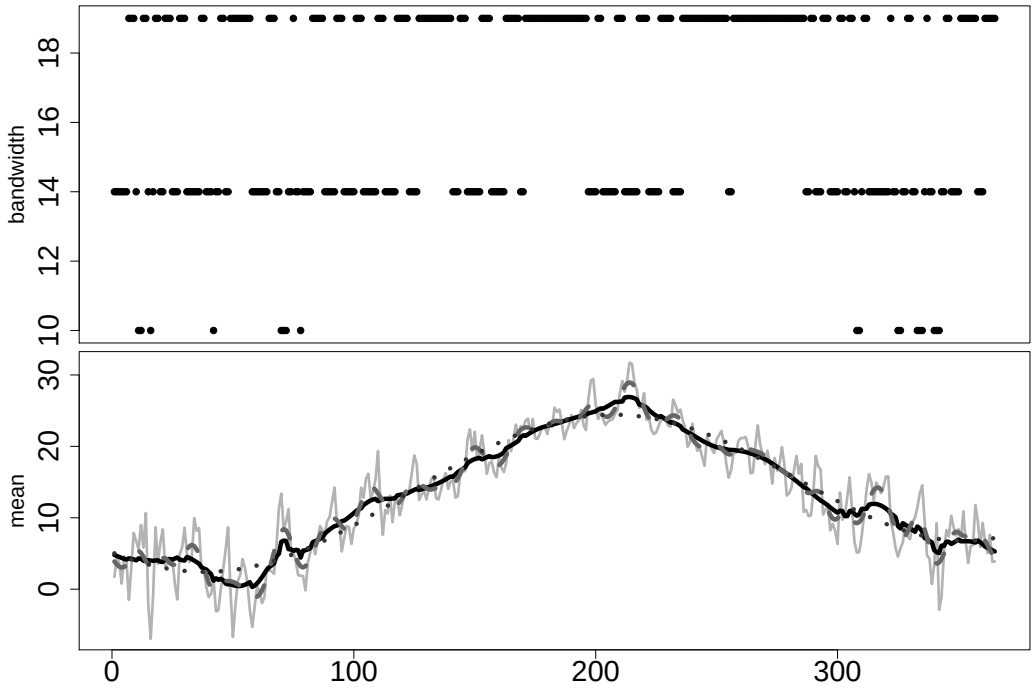


(a) Mean, 2008

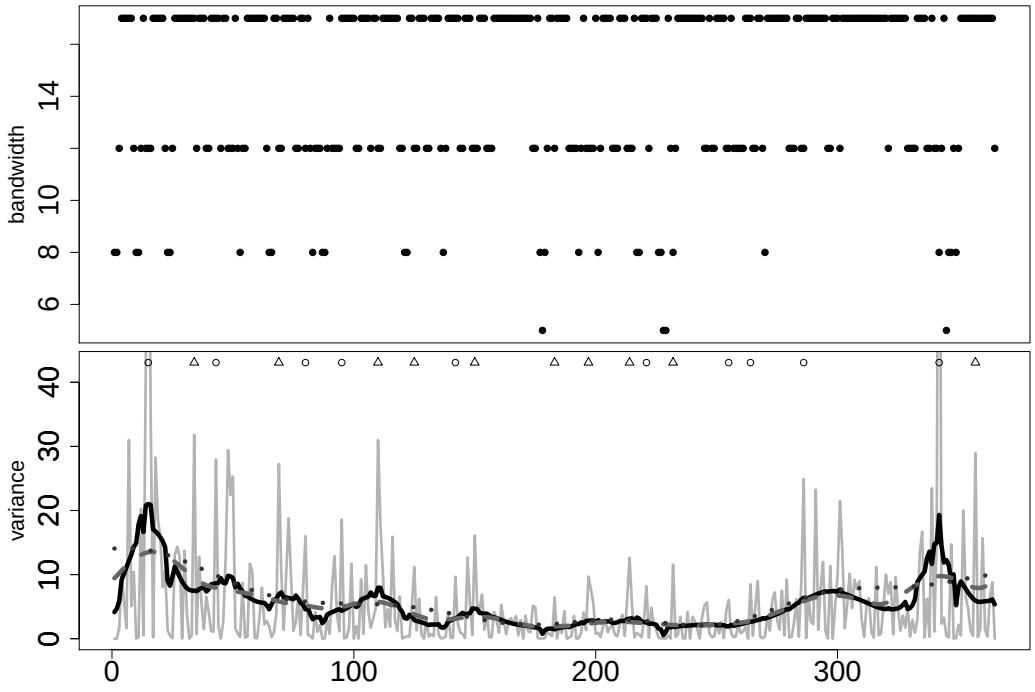


(b) Variance, 2008

Figure 6: Estimation of mean and variance for Kaohsiung. In both figure sequence of bandwidths (upper panel), averaged observations (solid grey line), nonparametric function estimation with fixed bandwidth (dashed grey line), adaptive bandwidth (solid black line) and truncated Fourier (dotted line) (bottom panel of each figure). Circles and triangles in each bottom panel for variance represents the 10 smallest and the 10 largest outliers respectively.

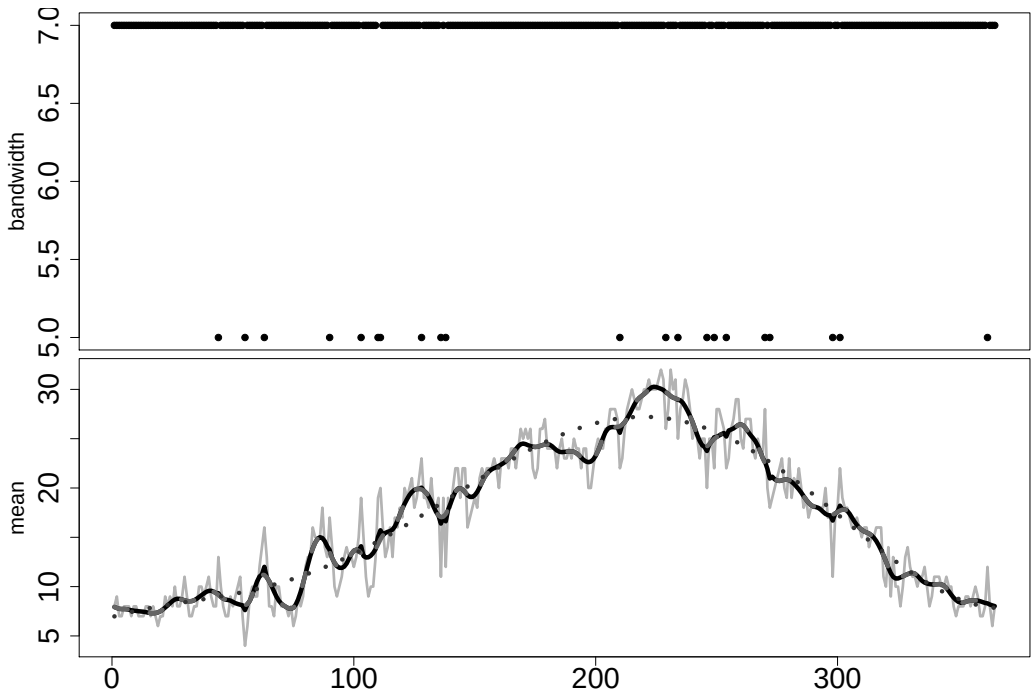


(a) Mean, 2007

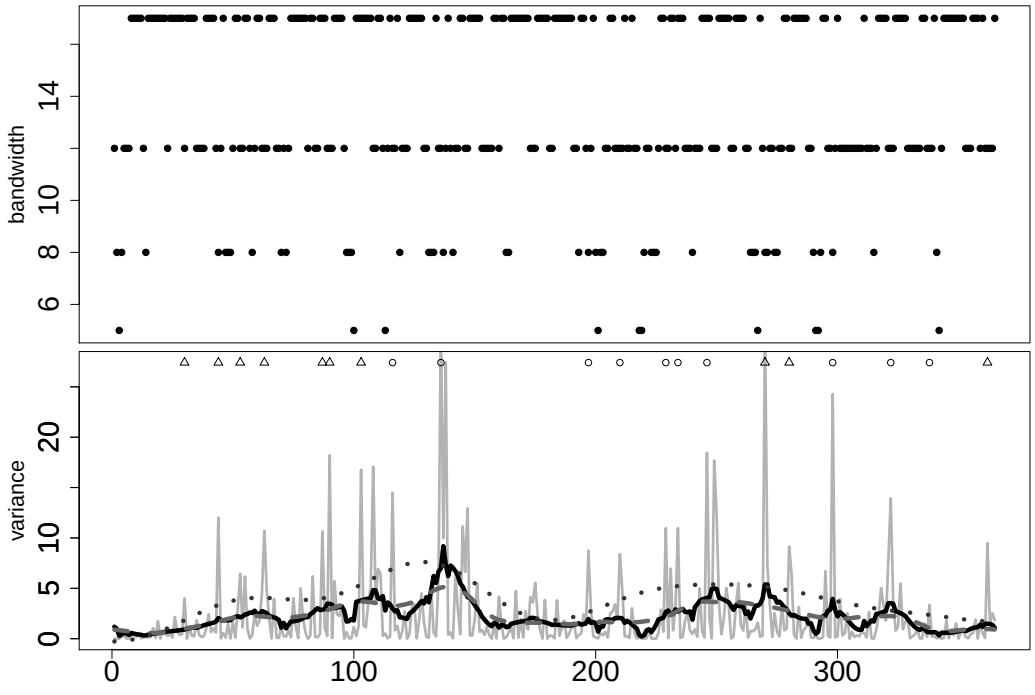


(b) Variance, 2007

Figure 7: Estimation of mean and variance for New York. In both figure sequence of bandwidths (upper panel), averaged observations (solid grey line), nonparametric function estimation with fixed bandwidth (dashed grey line), adaptive bandwidth (solid black line) and truncated Fourier (dotted line) (bottom panel of each figure). Circles and triangles in each bottom panel for variance represents the 10 smallest and the 10 largest outliers respectively.



(a) Mean, 2008



(b) Variance, 2008

Figure 8: Estimation of mean and variance for Tokyo. In both figure sequence of bandwidths (upper panel), averaged observations (solid grey line), nonparametric function estimation with fixed bandwidth (dashed grey line), adaptive bandwidth (solid black line) and truncated Fourier (dotted line) (bottom panel of each figure). Circles and triangles in each bottom panel for variance represents the 10 smallest and the 10 largest outliers respectively.

Method	Explanation
JoMe adMe adVa	Jointly estimated mean, adaptive bandwidth mean adaptive bandwidth variance
JoMe fiMe fiVa	Jointly estimated mean, fixed bandwidth mean fixed bandwidth variance
SeMe adMe adVa	Separately estimated mean, adaptive bandwidth mean adaptive bandwidth variance
SeMe fiMe fiVa	Separately estimated mean, fixed bandwidth mean fixed bandwidth variance
Locave	Aggregated approach with average of yearly empirical variance as the target
Locsep	Aggregated approach with each year's empirical variance as the target
Locmax	The optimal between Locave and Locsep (minimize the p value)
Fourier	Method with Fourier series fitting for mean and variance
CD	Method adopted by Campbell and Diebold (2005)

Table 3: Summary of methods

to ten dots of the upper and lower tails of the QQ-plots of square residuals respectively (see Figure 9 for the Berlin results). The top plots of Figures 5–8 show the mean case. Unlike the seasonal variance function, we do not observe a big variation of smoothness in the mean function. One can see that in all cities, the bandwidths vary over the yearly cycle with a slight degree of non homogeneity for Kaoshiung.

An approach to cope with the non normality brought in by more observations is to estimate mean functions year by year (SeMe), and then aggregate the residuals for variance estimation. We therefore estimate the joint/separate seasonal mean (JoMe/SeMe) and seasonal variance (Va) curves with a fixed bandwidth curve (fi) and an adaptive bandwidth curve (ad). (A summary of the estimation methods can be found in Table 3.) The average over years acts as a smoother when we consider more years. The estimated $AR(L)$ parameters for different cities using a joint/separate mean (JoMe/SeMe) with different bandwidth curves are illustrated in Table 4. The results again show that an $AR(3)$ fits the stylised facts of temperature well.

Kolmogorov–Smirnov (KS), Jarques–Bera (JB) and AD normality tests are taken to test the normality of the corrected residuals (after seasonal mean and variance). For each city, a rejection at 0.05 level is counted as 1 (else 0). The rejection rates over all the cities under different estimation techniques are displayed in Table 5. The results compare different periods ($1-5$ years) for the robustness of our methods. (Considering data histories longer than 5 years would not give us a better forecast performance and normality test results.)

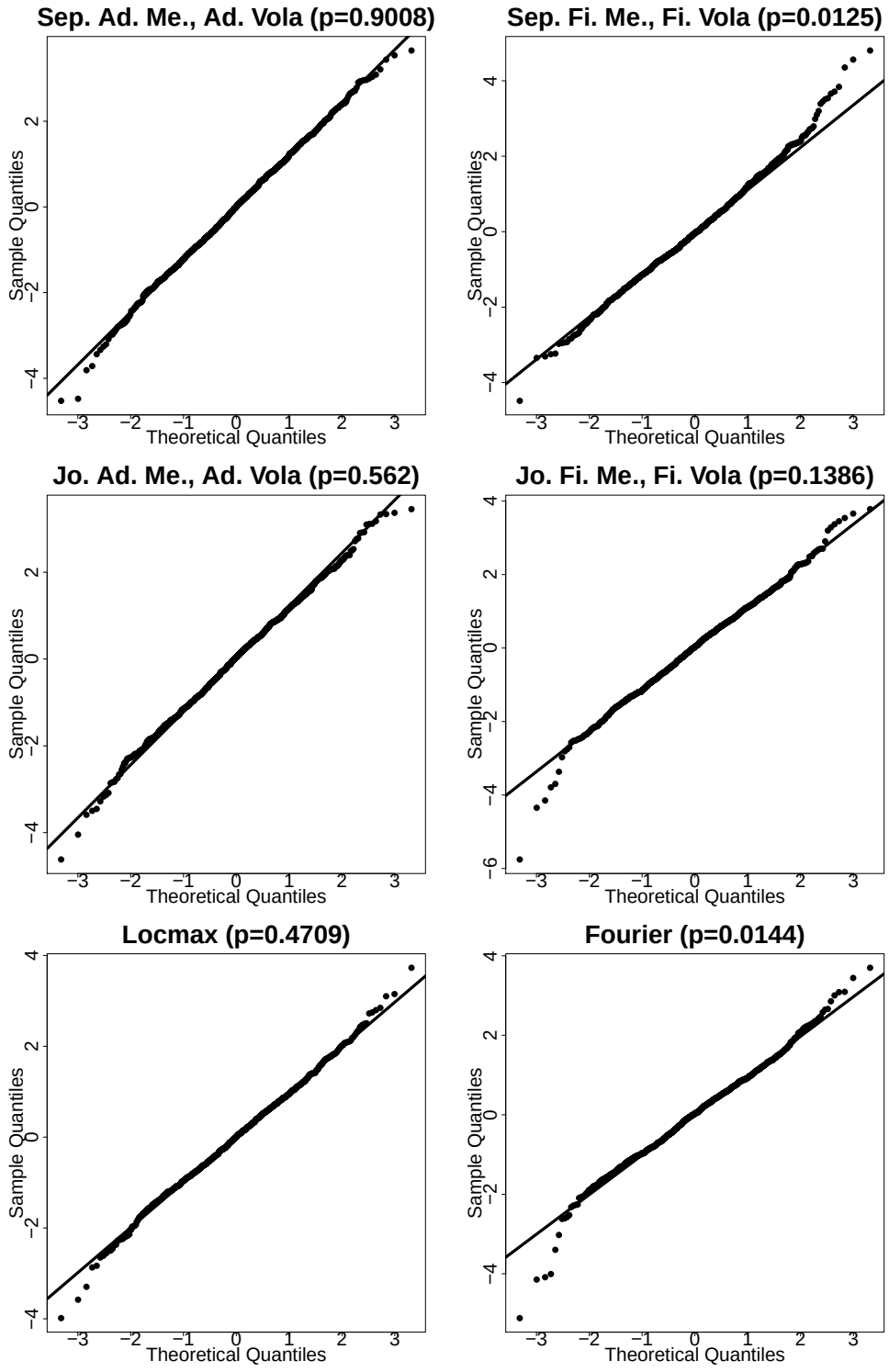


Figure 9: QQ-plot for standardised residuals from Berlin using different methods for the data from 2005-2007 (3 years). Please see Table 3 for a summary of methods.

		AR(p) parameters														CAR(p) parameters		
		1 year			2 years			3 years			4 years			5 years			2 years		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(1)$	$\hat{\alpha}_3(1)$
Berlin	JoMe adMe	0.744	-0.244	-0.008	0.972	-0.303	0.060	0.971	-0.286	0.072	0.960	-0.252	0.073	0.985	-0.288	0.101	2.015	1.318	0.202
	JoMe fiMe	0.301	-0.330	-0.267	0.908	-0.283	0.113	0.937	-0.277	0.113	0.939	-0.252	0.097	0.976	-0.289	0.114	2.024	1.337	0.199
	SeMe adMe adVo				0.795	-0.255	-0.057	0.807	-0.293	0.000	0.727	-0.248	-0.038	0.768	-0.290	0.000	2.204	1.664	0.517
	SeMe fiMe fiVo				0.344	-0.333	-0.196	0.346	-0.334	-0.176	0.294	-0.332	-0.175	0.326	-0.342	-0.165	2.655	2.644	1.185
	Locave				0.795	-0.255	-0.057	0.971	-0.286	-0.072	0.960	-0.252	-0.073	0.768	-0.290	0.000	2.204	1.664	0.517
	Locsep				0.795	-0.255	-0.057	0.971	-0.286	-0.072	0.960	-0.252	-0.073	0.768	-0.290	0.000	2.204	1.664	0.517
	Locmax				0.795	-0.255	-0.057	0.971	-0.286	-0.072	0.960	-0.252	-0.073	0.768	-0.290	0.000	2.204	1.664	0.517
	Fourier	0.968	-0.292	0.065	1.017	-0.309	0.075	1.017	-0.297	0.074	0.981	-0.264	0.079	1.000	-0.297	0.101	1.982	1.273	0.215
Kaohsiung	JoMe adMe	0.601	-0.132	-0.152	0.735	-0.105	-0.071	0.826	-0.204	0.000	0.807	-0.155	0.000	0.785	-0.139	0.000	2.215	1.569	0.354
	JoMe fiMe	0.507	-0.164	-0.200	0.725	-0.112	-0.080	0.806	-0.195	0.000	0.788	-0.146	0.000	0.773	-0.139	0.000	2.227	1.593	0.366
	SeMe adMe adVo				0.442	-0.138	-0,152	0.401	-0.200	-0.154	0.412	-0.159	-0.132	0.316	-0.194	-0.146	2.557	2.253	0.084
	SeMe fiMe adVo				0.504	-0.152	-0.210	0.502	-0.219	-0.206	0.479	-0.184	-0.196	0.457	-0.189	-0.190	2.495	2.143	0.857
	Locave				0.442	-0.138	-0.152	0.401	-0.200	-0.154	0.412	-0.159	-0.132	0.316	-0.194	-0.146	2.557	2.253	0.848
	Locsep				0.442	-0.138	-0.152	0.401	-0.200	-0.154	0.412	-0.159	-0.132	0.316	-0.194	-0.146	2.557	2.253	0.848
	Locmax				0.442	-0.138	-0.152	0.401	-0.200	-0.154	0.412	-0.159	-0.132	0.316	-0.194	-0.146	2.557	2.253	0.848
	Fourier	0.820	-0.116	-0.020	0.791	-0.102	-0.051	0.853	0.185	-0.000	0.820	-0.128	0.011	0.802	-0.126	0.001	2.208	1.520	0.363
New-York	JoMe adMe	0.601	-0.207	0.000	0.636	-0.197	0.161	0.728	-0.212	0.127	0.688	-0.212	0.165	0.708	-0.190	0.140	2.292	1.774	0.342
	JoMe fiMe	0.342	-0.277	-0.118	0.644	-0.205	0.138	0.692	-0.213	0.150	0.695	-0.222	0.158	0.711	-0.195	0.139	2.289	1.773	0.345
	SeMe adMe adVo				0.554	-0.247	0.000	0.495	-0.282	0.000	0.481	-0.287	0.000	0.475	-0.247	-0.050	2.289	1.773	0.345
	SeMe fiMe fiVo				0.278	-0.312	-0.118	0.283	-0.322	-0.133	0.299	-0.326	-0.120	0.288	-0.308	-0.152	2.712	2.732	1.172
	Locave				0.000	0.554	-0.247	0.495	-0.282	0.000	0.481	-0.287	0.000	0.475	0.247	-0.050	2.525	1.803	0.328
	Locsep				0.000	0.554	-0.247	0.495	-0.282	0.000	0.481	-0.287	0.000	0.475	0.247	-0.050	2.525	1.803	0.328
	Locmax				0.000	0.554	-0.247	0.495	-0.282	0.000	0.481	-0.287	0.000	0.475	0.247	-0.050	2.525	1.803	0.328
	Fourier	0.720	-0.184	0.066	0.763	-0.193	0.117	0.751	-0.205	0.134	0.750	-0.212	0.148	0.756	-0.190	0.128	2.244	1.678	0.306
Tokyo	JoMe adMe	0.152	-0.157	-0.294	0.431	-0.074	0.000	0.492	-0.087	0.000	0.541	-0.092	0.072	0.567	-0.102	0.083	2.433	1.968	0.452
	JoMe fiMe	0.158	-0.150	-0.296	0.452	-0.074	-0.054	0.512	-0.103	0.000	0.564	-0.097	0.045	0.575	0.105	0.072	2.425	1.745	0.248
	SeMe adMe adVo				0.360	-0.106	-0.162	0.330	-0.177	-0.136	0.307	-0.190	-0.138	0.317	-0.198	-0.124	2.639	2.384	0.907
	SeMe fiMe fiVo				0.225	-0.177	-0.245	0.265	-0.211	-0.184	0.269	-0.212	-0.171	0.252	-0.232	-0.171	2.774	2.726	1.197
	Locave				0.360	-0.106	-0.162	0.333	-0.177	-0.136	0.307	-0.190	-0.138	0.317	-0.198	-0.124	2.639	2.384	0.907
	Locsep				0.360	-0.106	-0.162	0.333	-0.177	-0.136	0.307	-0.190	-0.138	0.317	-0.198	-0.124	2.639	2.384	0.907
	Locmax				0.360	-0.106	-0.162	0.333	-0.177	-0.136	0.307	-0.190	-0.138	0.317	-0.198	-0.124	2.639	2.384	0.907
	Fourier	0.534	0.038	-0.039	0.561	-0.039	-0.017	0.581	-0.091	0.044	0.597	-0.088	0.060	0.615	-0.096	0.079	2.438	1.916	0.495

Table 4: $AR(L)$ parameters for Berlin, Kaohsiung, New York, Tokyo using joint/separate mean (JoMe/SeMe) with fixed bandwidth curve (fi), adaptive bandwidth curve (ad) seasonal mean (Me)curve. $CAR(L)$ parameters estimated with the last 2 years.

	Method	KS	JB	AD		KS	JB	AD
1 Year	JoMe adMe adVa	*0.000*	*0.174*	*0.164*				
	JoMe fiMe fiVa	0.006	0.200	0.270				
	Fourier	0.049	0.378	0.327				
	CD	0.086	0.499	0.426				
2 Years	JoMe adMe adVa	1.000	0.224	0.202	3 Years	0.968	0.354	0.367
	JoMe fiMe fiVa	0.998	0.431	0.390		0.869	0.571	0.533
	SeMe adMe adVa	1.000	0.073	0.043		1.000	*0.044*	*0.072*
	SeMe fiMe fiVa	1.000	0.305	0.261		0.976	0.358	0.367
	Locave	*0.001*	0.057	0.072		*0.004*	0.082	0.118
	Locsep	*0.001*	0.057	0.072		*0.004*	0.082	0.118
	Locmax	0.010	*0.051*	*0.034*		0.024	0.074	0.080
	Fourier	0.109	0.516	0.472		0.180	0.685	0.599
	CD	1.000	0.715	0.642		1.000	0.828	0.769
4 Years	JoMe adMe adVa	0.767	0.480	0.478	5 Years	0.585	0.547	0.539
	JoMe fiMe fiVa	0.608	0.646	0.618		0.483	0.702	0.669
	SeMe adMe adVa	0.975	*0.064*	*0.090*		0.747	*0.081*	*0.124*
	SeMe fiMe fiVa	0.778	0.468	0.433		0.463	0.546	0.506
	Locave	*0.007*	0.129	0.155		*0.009*	0.174	0.210
	Locsep	*0.007*	0.129	0.155		*0.009*	0.174	0.210
	Locmax	0.029	0.135	0.111		0.031	0.157	0.145
	Fourier	0.256	0.766	0.677		0.305	0.816	0.740
	CD	1.000	0.880	0.801		1.000	0.916	0.832

Table 5: Rejection rates of the normality at 5% level for 1000 cities with different history, methods of estimation and normality tests. Tests for normality are Kolmogorov–Smirnov (KS), Jarque–Bera (JB) and AD. Methods used: joint/separate mean (JoMe/SeMe) with fixed/adaptive (fi/ad) bandwidth for the mean/variance (Me/Va), Locave, Locsep, Locmax, truncated Fourier (Fourier) and CD model. Highlighted in italic are models with the smallest rejection rate for each goodness-of-fit (GoF) test and each history.

A higher rejection rate would indicate a poorer performance of the relevant method. To make our conclusion more general, we add 988 more cities, which are selected all around the world resulting in a total of 1000 cities, see Figures 10-11. The additional data is taken from NCDC Climate Data Online from $2007 - 2012$. We observe a superior performance of adaptive methods over the CD method and a truncated Fourier. The JoMe adMe adVa method with one year of history leads to the rejection rate up to of 0.174 which is more than twice smaller than using other methods. Considering more years of history, the rejection rate of the CD method comes close to 1.0 based on all tests and the rejection rate for the truncated Fourier based on the KS test is around 0.25 and based on two other tests, approaches 0.8. In contrary to CD and the truncated Fourier, rejection rates from all the adaptive methods are below 0.2 for all three tests. Moreover, one observes the rejection rate below 0.01 for the KS test for all years of history using the Locave and Locsep methods. SeMe adMe adVa method keeps the rejection rate for 3-5 years of history and JB and AD tests below 0.13. The Locmax procedure has a very stable performance over all the tests and all the history, with rejection rates being bounded by 0.16. Maps with marked locations on which the analysis has been performed using the period of five years of history and most conservative tests namely JB and AD are presented in Figures 10-11. Cities marked in blue are those, where the normality at a 5% level cannot be rejected using JB (Figure 10) and AD (Figure 11) tests, otherwise cities are marked in red. One clearly sees dominance of blue marked cities in the Locmax method (in both figures top left map) and the dominance of red marked cities in the other subplots. More detailed results for only 12 original cities can be found in the Supplementary Material.

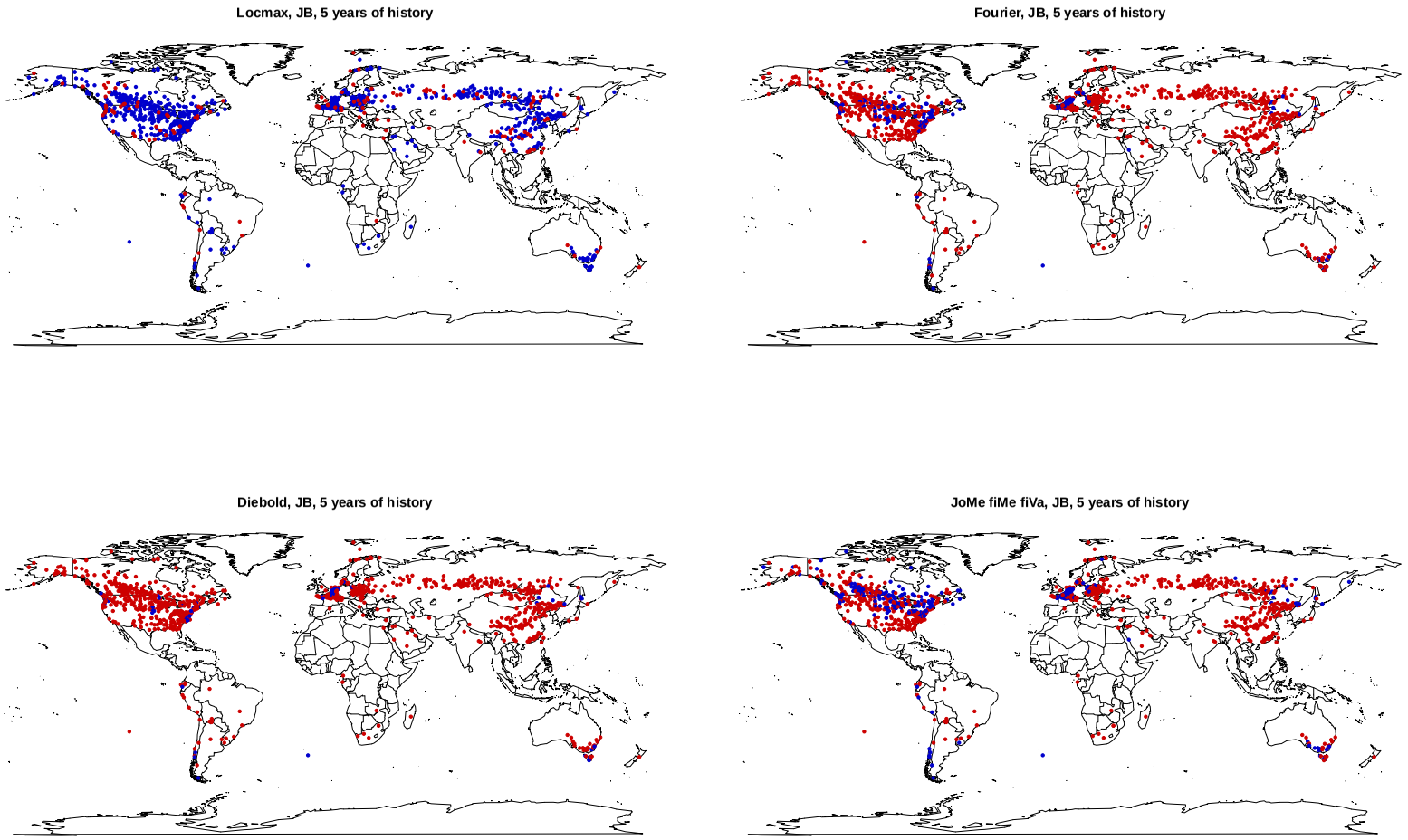


Figure 10: Map of locations where temperature are collected over period 2007-2011. Cities marked in blue do not reject the normality at 5% using the JB test and cities marked red do reject the normality hypothesis. In a clockwise direction, used methods are Locmax, truncated Fourier, CD and JoMe fiMe fiVa.

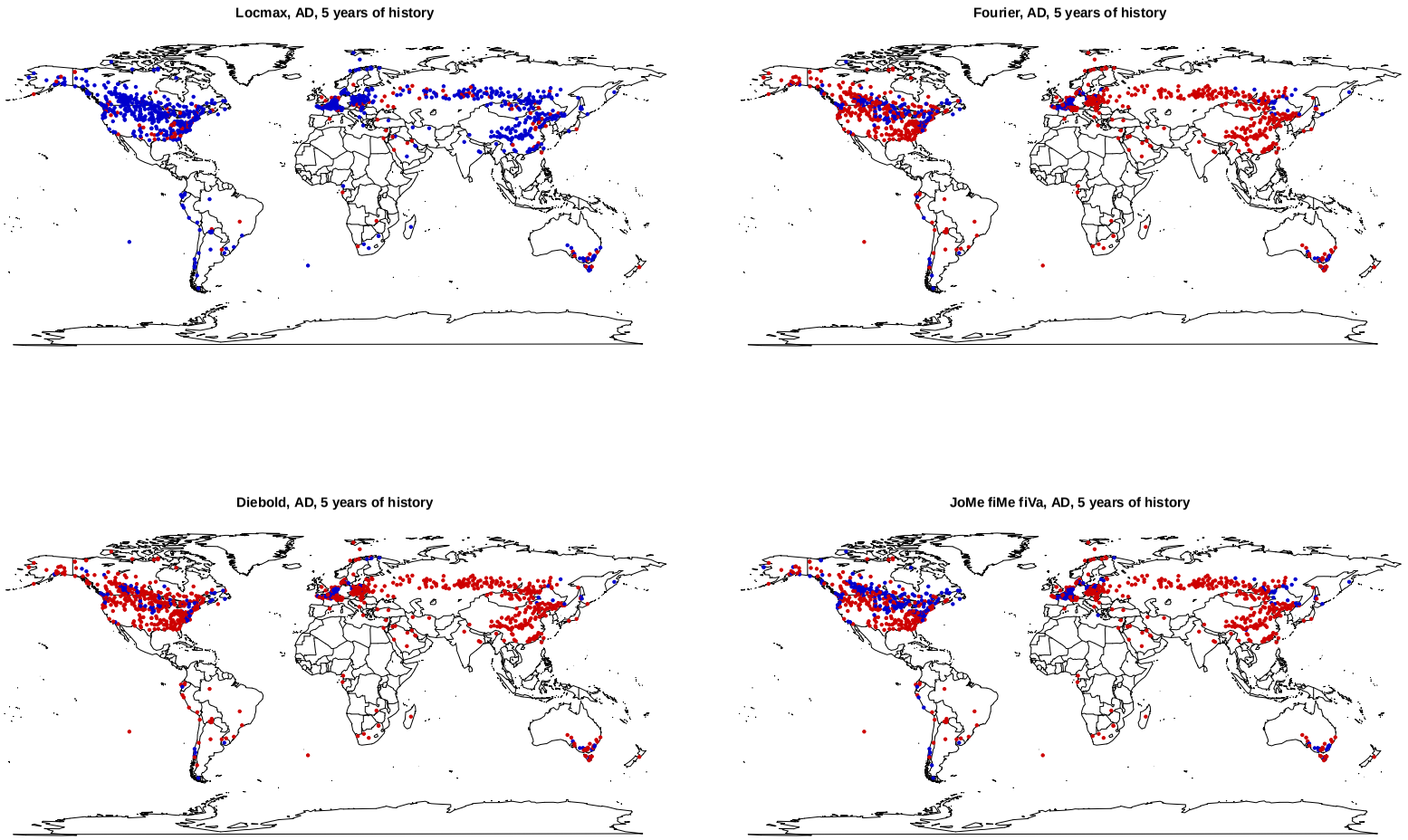


Figure 11: Map of locations where temperature are collected over period 2007-2011. Cities marked in blue do not reject the normality at 5% using the AD test and cities marked red do reject the normality hypothesis. In a clockwise direction, used methods are Locmax, truncated Fourier, CD and JoMe fiMe fiVa.

4 Forecast and comparison

In this section we compare the forecasting accuracy of the proposed models to the CD model. CD mentions that their point forecasts are always at least as good as the persistence and climatological forecasts, although not so good as the judgementally adjusted NWP forecast produced by EarthSat for a horizon of eight days. Therefore, a good performance of the technique presented here could potentially suggest that our time series model is relevant for weather derivatives.

In Figures 12 and 13 we compare the out-of-sample forecast performance between five methods, namely SeMe adMe adVo, Locmax, JoMe adMe adVo, truncated Fourier and CD. The comparison is provided at different time horizons $(1, \ldots, 150$ days) for Berlin, Kaohsiung, New York and Tokyo using 2 (Figure 12) and 3 (Figure 13) years of history. These figures contain information both on point forecast and interval forecast. The top panel of each plot shows the absolute deviation of the forecasted temperature from the true one, averaged over 10000 simulation paths. This may be considered as the quality of the point forecast. In these terms, as we see in most cities and over all time horizons, we have at least one localising method better than the CD method. The lower panel of each plot shows the averaged width of the point-wise confidence interval based on 10000 sample paths. These curves represent the efficiency of the models. Although the truncated Fourier series method also looks quite competitive in the point forecast, it usually has a very wide confidence interval, which is a sign of low efficiency. Other methods in this context are strictly better. The middle panel shows the coverage of the true temperature by the confidence interval, where larger values represent higher quality. In terms of interval forecast, we can see that from Figure 12 and 13 for most cities, we have at least one model which has better coverage with moderate width of confidence intervals. Moreover, we do not see outperforming behavior of the CD method over proposed adaptive techniques in almost all 12 cities. As a conclusion, we do not claim strict superiority over the CD method in forecasting, but conclude, that both methods are

quite competitive.

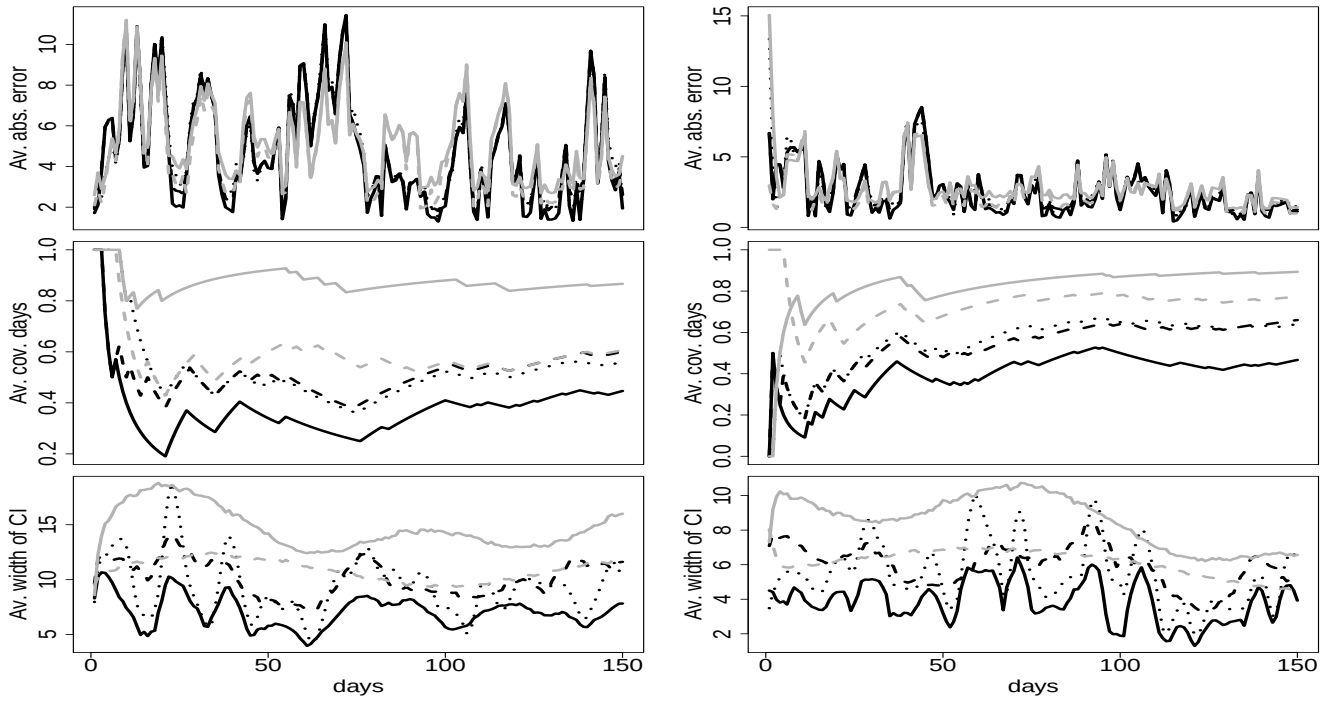
5 A temperature pricing example

Based on a model for the daily temperature evolution, futures and European options written on temperature indices traded at the Chicago Mercantile Exchange (CME) can be calibrated. Temperature futures are contracts written on different temperature indices measured over specified periods $[\tau_1, \tau_2]$ such as weeks, months, or quarters of a year. Temperature futures allow one party to profit if the realized index value is greater than a predetermined strike level and the other party benefits if the index value is below. The owner of a call (put) option written on futures $F(t, \tau_1, \tau_2)$ with exercise time $t \leq \tau_1$ and measurement period $[\tau_1, \tau_2]$ will receive $\max\{F(t, \tau_1, \tau_2) - K, 0\}$ ($\max\{K - F(t, \tau_1, \tau_2), 0\}$), where K denotes the strike level. In other words, in exchange for the payment of the premium, the call (put) option gives the buyer a payoff based upon the difference between the realized index value and the strike level.

The most common temperature indices $I(\tau_1, \tau_2)$ are: Heating Degree Day (HDD), Cooling Degree Day (CDD), Cumulative Averages Temperatures (CAT) or Average Accumulative Temperatures (AAT). The CAT index takes the accumulated average temperature over $[\tau_1, \tau_2]$:

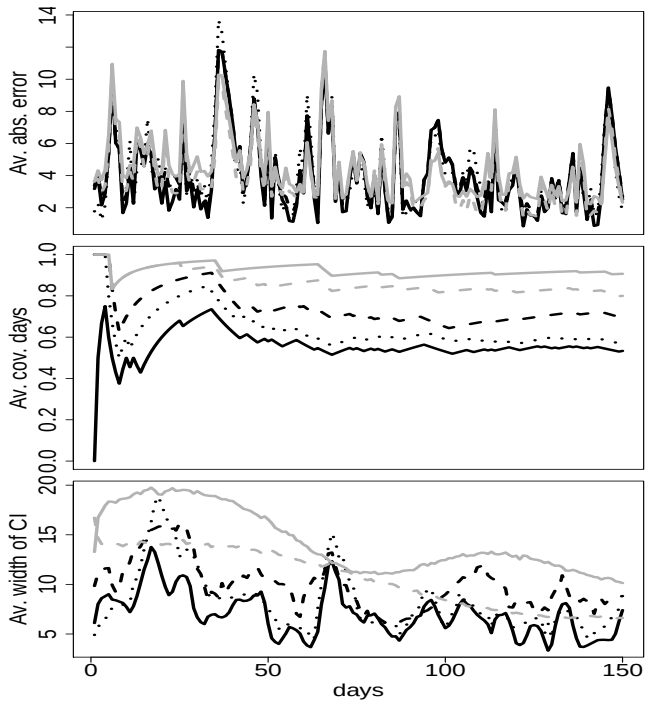
$$CAT(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} T_u du,$$

where $T_u = (T_{u,max} + T_{u,min})/2$ denotes the daily average temperature. The measurement period is usually defined in months or season. The HDD index measures the cumulative amount of average temperature below a threshold (typically 18°C or 65°F) over a period $[\tau_1, \tau_2]$: $\int_{\tau_1}^{\tau_2} \max(c - T_u, 0) du$. Similarly, the CDD index accumulates $\max(T_u - c, 0)$. At CME, CAT/CDD futures are traded for European cities, CDD/HDD for US, Canadian, and

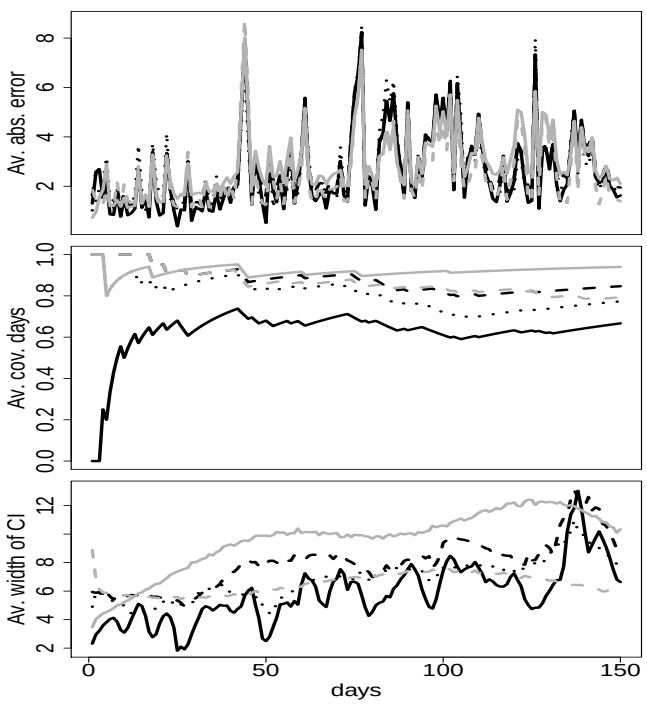


(a) Berlin (2006-2007)

(b) Kaohsiung (2007-2008)

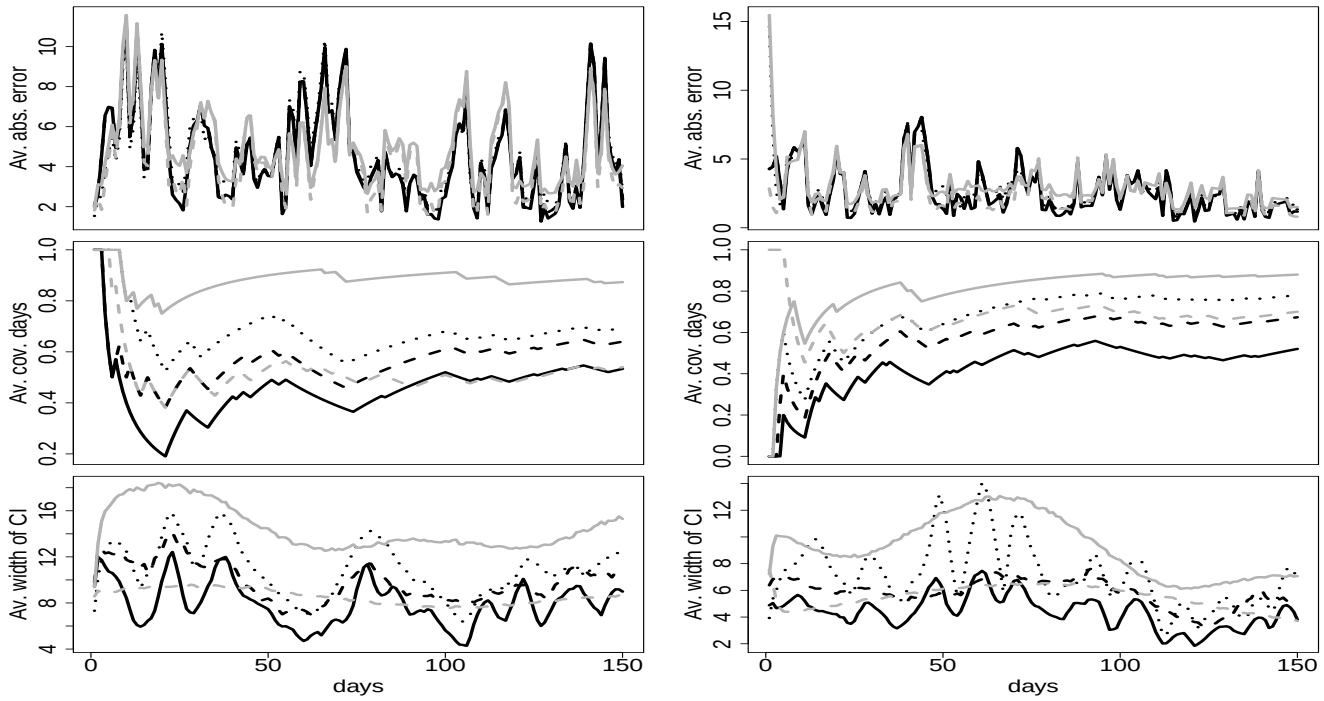


(c) New York (2006-2007)



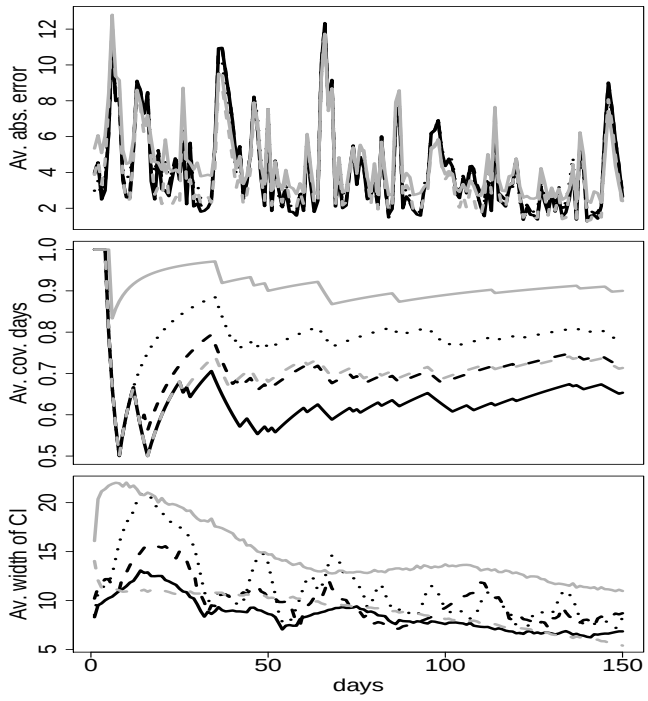
(d) Tokyo (2007-2008)

Figure 12: $h = 1, \ldots, 150$ days (X axis) ahead forecast for Berlin, Kaohsiung, New York and Tokyo (left to right, top to bottom); averaged absolute error (Y axis, upper panel), averaged coverage days (Y axis, middle panel), averaged width of the confidence 95% intervals (Y axis, lower panel), SeMe adMe adVo (solid black), Locmax (dashed grey), JoMe adMe adVo (dotted black), truncated Fourier (solid grey), CD (dashed black), fitted using 2 years of historical data and 10000 samples.

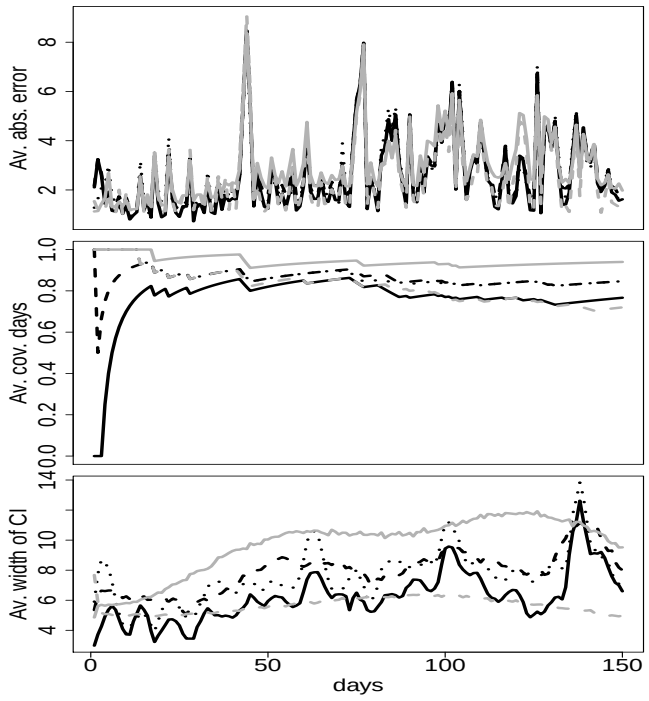


(a) Berlin (2005-2007)

(b) Kaohsiung (2006-2008)



(c) New York (2005-2007)



(d) Tokyo (2006-2008)

Figure 13: $h = 1, \ldots, 150$ days (X axis) ahead forecast for Berlin, Kaohsiung, New York and Tokyo (left to right, top to bottom); averaged absolute error (Y axis, upper panel), averaged coverage days (Y axis, middle panel), averaged width of the confidence 95% intervals (Y axis, lower panel), SeMe adMe adVo (solid black), Locmax (dashed grey), JoMe adMe adVo (dotted black), truncated Fourier (solid grey), CD (dashed black), fitted using 3 years of historical data and 10000 samples.

Australian cities, and AAT for Japanese cities. Note that these temperature indices are the underlying and not the temperature itself. The options at CME are cash settled, i.e., the owner of a future receives 20 times the Degree Day Index at the end of the measurement period, in return for a fixed price. At time t, CME trades different contracts $i = 1, \ldots, N$ with measurement period $0 \leq t \leq \tau_1^i < \tau_2^i$ (usually the length between τ_1^i and τ_2^i is one month). For example, a contract with $i = 7$ is six months ahead from the trading day t. For the US and Europe CAT/CDD/HDD futures, N is usually equal to 7 (April–November or November–April), while for Asia, $N = 12$ (January–December).

Recall that we adopt the $CAR(L)$ model in (1) for the detrended temperature time series, and the autoregressive process $AR(L)$ in (2) can be seen as a discretely sampled continuous time process ($CAR(L)$) (1) driven by one dimensional Brownian motion. The detailed demonstration can be found in the Appendix 7.2.

The fact that temperature's random factor is close to the normal distribution, as disclosed in the analysis of the residuals before, motivates the use of a Brownian motion as the noise in the Ornstein-Uhlenbeck process. Moreover ACF-plots of the squared residuals presented in the Supplementary Material demonstrate the success of the localizing method to explain deterministic variations in temperature data. They do not show signs of stochastic volatility: the squared residuals do not have an exponentially decaying ACF. This contradicts results found in Benth and Benth (2011) and Benth and Benth (2012) and suggests to us that the non-Gaussian shocks found in the literature are the result of model mis-specification. The continuous analogue of the CD model is however difficult to estimate. Thus the model in (1) is simpler than CD's one and provides a better fit to the data.

The temperature futures price is the risk adjusted index, given today's filtration $\mathcal{F}_t$

$$F_I(t, \tau_1, \tau_2) = \mathsf{E}^Q \left[I(\tau_1, \tau_2) | \mathcal{F}_t \right], \tag{14}$$

with $I(\tau_1, \tau_2)$ being one of the indices CAT, HDD or CDD. The expectation is computed

under a risk neutral pricing probability Q and is equivalent to the physical measure P under which the discounted temperature index is a Q-martingale. To evaluate (14), we need to know the temperature index dynamics under Q. We restrict the class of pricing probabilities to those that can be parametrized via $Q = Q_\lambda$, where equivalent changes of measures are simply associated with changes of drift. Thus, in the modelling of the dynamics of futures prices written on temperature indices, it is natural to define a parameter measuring the market price of risk (MPR) λ_t, which can be calibrated from traded (CAT/CDD/HDD) derivative type contracts. The temperature dynamics in (1) under Q_λ become:

$$d\mathbf{X}_t = (\mathbf{A}\mathbf{X}_t + \mathbf{e}_L\sigma_t\lambda_t)dt + \mathbf{e}_L\sigma_t dB_t^\lambda, \tag{15}$$

where B_t^λ is a Brownian motion for any time before the end of the trading time and a martingale under Q_λ. Then, for $0 \leq t \leq \tau_1 < \tau_2$, the explicit form of an CAT futures price is given by:

$$\begin{aligned} F_{CAT}(t,\tau_1,\tau_2,\Lambda_t,\sigma_t,\lambda_t) &= \mathsf{E}^{Q_\lambda}\left[\int_{\tau_1}^{\tau_2} T_u du|\mathcal{F}_t\right] \\ &= \int_{\tau_1}^{\tau_2}\Lambda_u du + \mathbf{a}_{t,\tau_1,\tau_2}\mathbf{X}_t + \int_t^{\tau_1}\lambda_u\sigma_u\mathbf{a}_{t,\tau_1,\tau_2}\mathbf{e}_L du \\ &+ \int_{\tau_1}^{\tau_2}\lambda_u\sigma_u\mathbf{e}_1^\top\mathbf{A}^{-1}\left[\exp\left\{\mathbf{A}(\tau_2-u)\right\} - I_L\right]\mathbf{e}_L du, \end{aligned} \tag{16}$$

with $\mathbf{a}_{t,\tau_1,\tau_2} = \mathbf{e}_1^\top\mathbf{A}^{-1}\left[\exp\left\{\mathbf{A}(\tau_2-t)\right\} - \exp\left\{\mathbf{A}(\tau_1-t)\right\}\right]$ and I_L the $L\times L$ identity matrix. Similarly one can compute the price dynamics of CDD and HDD, see Benth et al. (2007). The CAR model (1) provides the analytical formula (16). Note that all constituents except λ_t in the left and right side of (16) are known or estimable (Λ_t and σ_t are out-of-sample estimates as in the previous section), hence the calibration of the MPR from market data turns out to be an inverse problem in terms of λ_t.

Assuming that the parametrization of the MPR is of a constant form for each observed contract ($\lambda_u = \lambda_{t,\tau_1^i,\tau_2^i}$ in (16) for $u \in [\tau_1,\tau_2]$), one can calibrate the MPR for every combina-

tion of (t, τ_1^i, τ_2^i), $i = 1, \ldots, N$ contracts, by inverting the pricing formulae in (16) with the observed CME market prices at time t, $(F_{t,i,CME})$ with respect to λ as:

$$\hat{\lambda}_{t,\tau_1^i,\tau_2^i} = \arg\min_{\lambda} |F_{CAT}(t, \tau_1^i, \tau_2^i, \hat{\Lambda}_t, \hat{\sigma}_t, \lambda) - F_{t,i,CME}|. \tag{17}$$

We name $\hat{\lambda}_{t,\tau_1^i,\tau_2^i}$ as implied MPR. For fixed time t, assuming that λ_t remains the same for different contracts with different maturities, to evaluate the estimation of $\hat{\lambda}_t$ for a particular contract i, the observed price $F_{t,i,CME}$ for this contract can be excluded for the estimation. We have then the cross validated estimation by leaving one contract out:

$$\hat{\lambda}_{t,\tau_1^i,\tau_2^i,CV} = \arg\min_{\lambda} \sum_{j=1;j\neq i}^{N} \{F_{CAT}(t, \tau_1^j, \tau_2^j, \hat{\Lambda}_t, \hat{\sigma}_t, \lambda) - F_{t,j,CME}\}^2. \tag{18}$$

Other specifications of the MPR for temperature derivatives have been explored in Härdle and López Cabrera (2012), where the authors argue that a constant MPR is sufficient for pricing purposes. This might be compared with complete markets, where the MPR is minus the Sharp ratio $(\mu_t - r)/\sigma_t^F$, where μ_t and σ_t^F denote the mean and standard deviation of traded futures, and r is the risk free interest rate. From now on, pricing follows (16) with an MPR from (17), (18) and with Λ_t and σ_t estimated via the localisation techniques.

Observe that calibrations in (17), (18) are only valid if a weather derivative market exists, like e.g. for Berlin and Tokyo. In order to price temperature derivatives for regions with no weather derivative markets, like Kaohsiung, one can use the implied MPR of traded futures of a neighbouring market, e.g. Tokyo AAT futures. Thus, by finding a relationship between the MPR and the seasonal variance one can use this as a proxy to price over the counter (OTC) AAT futures for Kaohsiung. This is acceptable since the stylized facts of temperature in Tokyo reveal similarities to that of Kaohsiung. However generally, we are aware of arbitrage opportunities across the two different markets, therefore this approach cannot be generalized for every second weather derivative markets. Considering that the MPR is a risk premium

per unit of volatility, one can project the implied MPR on the state variables related to volatility. An insight into Tokyo's AAT futures, which can be employed for the Kaohsiung case, can be achieved by regressing the averaged implied MPR (17) against the variation:

$$\hat{\lambda}_{\tau_1^i,\tau_2^i} = 4.08 - 2.19\hat{\sigma}^2_{\tau_1^i,\tau_2^i} + 0.28\hat{\sigma}^4_{\tau_1^i,\tau_2^i},$$

where $\hat{\lambda}_{\tau_1^i,\tau_2^i} \stackrel{\text{def}}{=} (\tau_2^i - \tau_1^i)^{-1} \sum_{t=\tau_1^i}^{\tau_2^i} \hat{\lambda}_{t,\tau_1^i,\tau_2^i}$, $\hat{\sigma}^2_{\tau_1^i,\tau_2^i} \stackrel{\text{def}}{=} (\tau_2^i - \tau_1^i)^{-1} \sum_{t=\tau_1^i}^{\tau_2^i} \hat{\sigma}_t^2$, $\hat{\sigma}^4_{\tau_1^i,\tau_2^i} \stackrel{\text{def}}{=} (\tau_2^i - \tau_1^i)^{-1} \sum_{t=\tau_1^i}^{\tau_2^i} \hat{\sigma}_t^4$ and $R^2_{adj} = 0.71$. Plugging the corresponding $\hat{\sigma}_t^2$, $\hat{\sigma}_t^4$ values for Kaohsiung into this equation let us price such a non CME traded weather derivative via (16).

We compare the prices obtained with localisation procedures ('localised' prices) for Λ_t and σ_t (SeMe adMe adVo (AdaptBW), Locmax) with prices estimated under fixed bandwidth (SeMe fiMe fiVo (FixedBW)) and truncated Fourier series.

To judge the performance of the models, we compute the root mean squared errors (RMSE) between the market prices $F_{t,i,CME}$ (benchmark) and the estimated out-of-sample model prices
$F_I(t, \tau_1^i, \tau_2^i, \hat{\Lambda}_t, \hat{\sigma}_t, \hat{\lambda}_{t-1,\tau_1^i,\tau_2^i,CV})$ $(i = 1, \ldots, N)$:

$$RMSE(\tau_1^i, \tau_2^i) \;=\; \sqrt{|\mathfrak{T}|^{-1} \sum_{t \in \mathfrak{T}} \left\{ F_I(t, \tau_1^i, \tau_2^i, \hat{\Lambda}_t, \hat{\sigma}_t, \hat{\lambda}_{t-1,\tau_1^i,\tau_2^i,CV}) - F_{t,i,CME} \right\}^2},$$

in Table 6, where $\mathfrak{T}$ is the set of days when the contract i with the measurement period (τ_1^i, τ_2^i) was traded. The results show smaller RMSE when futures prices are estimated via localisation techniques, which in general outperforms the prices based on the truncated Fourier series (Benth). This suggests that our calibrated MPR embeds information on the risk and uncertainty in the market, which is helpful in analyzing market risk. Also, as mentioned before, this information may help to price OTC derivatives in the same market.

These results provide insight on pricing related to the stylized facts (seasonal effect, inter-

temporal correlation, etc.) of weather data. The role of the terms in the CAT futures price formally confirms this. To illustrate this point, consider, for example, the purchase of a May CAT contract for Berlin on 20070427, which starts measurement at time $\tau_1 = 20070501$ and finished at $\tau_2 = 20070531$. Setting a constant MPR (for example $\lambda = 0.20$), the first term of (16) is equal to 431.060, the second, third and fourth terms lead to 11.531, 0.8690 and 13.5390 respectively. The seasonal effect in mean Λ_t plays an important role in the level of the futures price, as it explains almost 94% of the price which is 457. Observe that the seasonal volatility σ_t also contributes to the CAT futures price since it enters in the second term (hidden in $\mathbf{X}_t$) and in the last two terms of the CAT pricing formulae. Therefore, as we get closer to the measurement period, temperature variations given by the seasonal variance (σ_t^2) will contribute to the futures prices and clearly display the Samuelson effect that is typical in mean-reverting markets: at any given time, seasonal volatility decreases with time to delivery.

6 Conclusions

We show that temperature risk stochastics are closer to Gaussian when applying adaptive statistical methods for seasonal mean and seasonal variance. This suggests to us that the non-Gaussian shocks found in the literature are truly a result of misspecification. We found that the localisation method performs well, and it is robust to the specification given for Λ_t or σ_t. Moreover intertemporal correlations demonstrate the success of the localizing methods to explain deterministic variations in temperature data. We also observed that the proposed method outperforms the standard estimation methods in most of the cases. Our results provide important insights into how weather is priced at the CME and how the observed prices conform with the stylized facts of weather data. Finally, our adaptive technique on localising temperature risk is both an excellent temperature modeling tool as well as a novel and more market driven pricer.

			RMSE between models' prices and F_{CME}			
				$\hat{\lambda}_{t-1,CV}$		
Type	MP	n	AdaptBW	FixedBW	Locmax	Fourier
Berlin-CAT	200704	230	2.868	*2.617*	2.876	9.665
Berlin-CAT	200705	6	79.802	84.078	*79.169*	126.8
Berlin-CAT	200706	58	*2.033*	3.078	2.662	68.262
Berlin-CAT	200707	79	*31.774*	46.633	32.565	45.125
Berlin-CAT	200709	121	*25.17*	39.337	25.485	26.773
Essen-CAT	200804	74	*75.676*	75.686	*75.676*	76.519
Essen-CAT	200805	100	21.871	*21.845*	21.871	22.628
Essen-CAT	200806	79	7.225	*7.131*	7.225	19.15
Essen-CAT	200807	140	*59.392*	59.47	*59.392*	62.318
Essen-CAT	200808	164	*73.511*	73.548	*73.511*	74.469
Essen-CAT	200809	181	6.885	*6.837*	6.885	12.932
London-CAT	200805	100	43.06	*32.377*	40.505	58.495
London-CAT	200806	40	*1.461*	2.56	2.709	6.063
London-CAT	200807	142	*2.467*	2.824	4.745	9.81
London-CAT	200808	163	27.333	27.204	*26.88*	31.23
London-CAT	200809	184	*36.201*	37.255	37.941	41.861
Tokyo-AAT	200903	18	4.922	*1.354*	8.418	26.287
Tokyo-AAT	200904	18	*28.967*	29.401	56.975	76.489
Tokyo-AAT	200905	18	58.553	*54.353*	90.269	77.8
Tokyo-AAT	200906	18	*49.993*	52.228	52.678	*16.35*
Tokyo-AAT	200907	18	24.093	27.72	*21.954*	42.34

Table 6: RMSE between the weather futures listed at CME and estimated weather futures $F_I(t, \tau_1^i, \tau_2^i, \hat{\Lambda}_t, \hat{\sigma}_t, \hat{\lambda}_{t-1})$ with $\hat{\lambda}_{t-1} = \hat{\lambda}_{t-1,CV}$. τ_1^i and τ_2^i are the first and the last day of the measurement period (MP, yyyymm) respectively. Prices are estimated under different estimations schemes ($\hat{\Lambda}_t, \hat{\sigma}_t$ under AdaptBW, FixedBW, Locmax and truncated Fourier). n corresponds to the number of trading days for a given MP.

7 Appendix

7.1 Theoretical Background

We now briefly introduce the theoretical background for the adaptation procedure. For $\ell < k$, the accuracy of the estimation is measured by the fitted likelihood ratio (LR):

$$L(W^\ell, \tilde{\theta}_\ell, \tilde{\theta}_k) \stackrel{\text{def}}{=} L(W^\ell, \tilde{\theta}_\ell) - L(W^\ell, \tilde{\theta}_k). \tag{19}$$

For the Gaussian risk factor situation the variance σ_t^2 (or trend Λ_t) estimation is carried out within an exponential family framework, so the LR can be written in a closed form:

$$\begin{aligned} L(W^k, \tilde{\theta}_k, \theta^*) &\stackrel{\text{def}}{=} N_k \mathcal{K}(\tilde{\theta}_k, \theta^*) \\ &= -\{\log(\tilde{\theta}_k/\theta^*) + 1 - \theta^*/\tilde{\theta}_k\}/2, \end{aligned} \tag{20}$$

where $N_k = J \sum_{t=1}^{365} w(s,t,h_k)$ and $\mathcal{K}(\tilde{\theta}_k, \theta^*)$ is the Kullback-Leibler divergence (21) between two normal distributions with variances $\tilde{\theta}_k$ and θ^*. Note that (20) is the divergence for exactly this case. For trend Λ_t estimation, it has to be replaced by $(\tilde{\theta}_k - \theta^*)^2/(2\sigma^2)$.

Recall that the Kullback-Leibler divergence of two distributions with densities $p(x)$ and $q(x)$ is

$$\mathcal{K}\{p(x), q(x)\} \stackrel{\text{def}}{=} \mathsf{E}_{p(.)} \log \frac{p(x)}{q(x)} \quad . \tag{21}$$

To guarantee the feasibility of the tests, we need moment bounds and confidence sets for the LR that will guarantee that the MLE is concentrated in the level set of the likelihood ratio process (indexed by the number of observations) around the true parameter, see Polzehl and Spokoiny (2006) and Mercurio and Spokoiny (2004). Below we state a result along this line for the variance (a similar bound can be derived for the mean).

Theorem 7.1 *[Spokoiny (2009)] Assuming that $\theta(t) = \theta^*$ for any $t \in [1, 365]$, then for $\mathfrak{z} > 0$ and $k \in 1, \ldots, K, r > 0$, denote by $\mathrm{P}_{\theta^*}(.)$ the measure corresponding to (9). We obtain*

$$\mathrm{P}_{\theta^*}\left\{L(W^k, \tilde{\theta}_k, \theta^*) > \mathfrak{z}\right\} \leq 2\exp\left(-\mathfrak{z}\right) \tag{22}$$

and a risk bound for a power loss function:

$$\mathsf{E}_{\theta^*}|L(W^k, \tilde{\theta}_k, \theta^*)|^r \leq \mathfrak{r}_r, \tag{23}$$

where $\mathfrak{r}_r = 2r\int_{\mathfrak{z}\geq 0}\mathfrak{z}^{r-1}\exp(-\mathfrak{z})d\mathfrak{z}$. This polynomial bound applies to all localising schemes W^k simultaneously.

The risk bound (23) allows us to define likelihood based confidence sets since together with (22) it tells us that the likelihood process is stochastically bounded. The confidence sets are therefore defined with critical values $\mathfrak{z}_k$ to level α as shown in (10).

The LMS algorithm is illustrated in Figure 14. For every estimate $\tilde{\theta}_k$ the corresponding confidence set is shown. If the horizontal line originating in $\tilde{\theta}_k$ does not cross all the preceding intervals then the selection algorithm terminates.

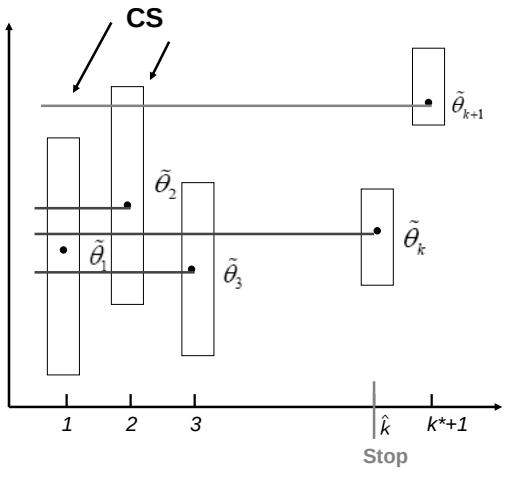


Figure 14: Illustration of the LMS

A further integrated approach is to consider an iterative algorithm, which iterates between

estimating the seasonal component and the variance $\theta(t) = \{\Lambda_t, \sigma_t^2\}$. This algorithm can further cope with heteroscedasticity in the corrected residuals after seasonality in mean and variance components. The procedure is:

Step 1. Estimate $\hat{\beta}$ in an initial Λ_t^0 using a truncated Fourier series or any other deterministic function;

Step 2. For fixed $\hat{\Lambda}_{s,\nu} = \{\hat{\Lambda}'_{s,\nu}, \hat{\Lambda}''_{s,\nu}\}^\top, s = \{1, \ldots, 365\}$ from last step ν, and fixed $\hat{\beta}$, get $\hat{\sigma}^2_{s,\nu+1}$ by

$$\begin{aligned}\hat{\sigma}^2_{s,\nu+1} &= \arg\min_{\sigma^2} \sum_{t=1}^{365} \sum_{j=0}^{J} [\{T_{365j+t} - \hat{\Lambda}'_{s,\nu} - \hat{\Lambda}''_{s,\nu}(t-s) \\ &- \sum_{l=1}^{L} \hat{\beta}_l X_{365j+t-l}\}^2/2\sigma^2 + \log(2\pi\sigma^2)/2] w(s,t,h'_k);\end{aligned}$$

Step 3. For fixed $\hat{\sigma}^2_{s,\nu+1}$ and $\hat{\beta}$, we estimate $\hat{\Lambda}_{s,\nu+1}, s = \{1, \ldots, 365\}$ via another local adaptive procedure:

$$\hat{\Lambda}_{s,\nu+1} = \arg\min_{\{\Lambda', \Lambda''\}^\top} \sum_{t=1}^{365} \sum_{j=0}^{J} \left\{ T_{365j+t} - \Lambda' - \Lambda''(t-s) - \sum_{l=1}^{L} \hat{\beta}_l X_{365j+t-l} \right\}^2 w(s,t,h'_k)/2\hat{\sigma}^2_{s,\nu+1},$$

where $\{h'_1, h'_2, h'_3, \ldots, h'_{K'}\}$ is a sequence of bandwidths;

Step 4. Repeat steps 2 and 3 until both $|\hat{\Lambda}_{t,\nu+1} - \hat{\Lambda}_{t,\nu}| < \pi_1$ and $|\hat{\sigma}^2_{t,\nu+1} - \hat{\sigma}^2_{t,\nu}| < \pi_2$ for some constants π_1 and π_2.

Our empirical implementation suggests that one iteration is enough. The LMS methods require Critical Values $\mathfrak{z}_k$, which define the significance for the LR statistics $L(W^\ell, \tilde{\theta}_\ell, \tilde{\theta}_k)$ or alternatively the length of the confidence interval (see (22)) at each step. As can be seen from above, the Critical Values are calibrated from the "propagation condition" below which ensures a desired level of type one error. To be more specific, for every step k, define $\hat{\theta}_k$ as the "survived estimator" after the kth step (if the estimator is not rejected up to step k, then

$\hat{\theta}_k = \tilde{\theta}_k$, else if the estimator has been rejected at step $l < k$, then $\hat{\theta}_k = \tilde{\theta}_l$). Measure the closeness of $\tilde{\theta}_k$ and $\hat{\theta}_k$ by

$$\mathsf{E}_{\theta^*} |L(W^k, \tilde{\theta}_k, \hat{\theta}_k)|^r \leq \alpha \mathfrak{r}_r \tag{24}$$

for $k = 1, \ldots, K$ with $\mathfrak{r}_r$ the parametric risk bound in (23) and α a control parameter corresponding to the type one error. In fact

$$\mathsf{E}_{\theta^*} |L(W^k, \tilde{\theta}_k, \hat{\theta}_k)|^r \to \mathrm{P}_{\theta^*}(\tilde{\theta}_k \neq \hat{\theta}_k)$$

for $r \to 0$, therefore α can be interpreted as a false alarm probability.

More precisely, if step k is accepted as described in Figure 14, then $\tilde{\theta}_k = \hat{\theta}_k$ and a non-zero loss $\mathsf{E}_{\theta^*} L(W^k, \tilde{\theta}_k, \hat{\theta}_k)$ can only occur if the estimator has been rejected before or at step k, which under the homogeneous parametric model case, is denoted as a "false alarm".

A risk bound for a constant model ($\theta(t) = \theta^*$) has been given in (24). In order to expand this to a nonparametric $\theta(t)$, the "Small Modeling Bias (SMB)" condition is employed:

$$\Delta(\theta) \stackrel{\text{def}}{=} \sum_{t=1}^{365} \mathcal{K}\{\theta(t), \theta\} \, \mathbf{I}\{w(s, t, h_k) > 0\} \leq \Delta, \forall k < k^*, \tag{25}$$

where k^* is the maximum k satisfying (25), also called "oracle". Consequently the estimation risk for $\theta(t)$ is described for $k \leq k^*$ by the "propagation" property:

$$\mathsf{E}_{\theta(\cdot)} \log\{1 + |L(W^k, \tilde{\theta}_k, \hat{\theta}_k)|^r / \mathfrak{r}_r\} \leq \Delta + \alpha. \tag{26}$$

An estimate for the oracle k^* is given via the adaptive estimate $\hat{\theta}_{\hat{k}}$. The estimate $\hat{\theta}_{\hat{k}}$ behaves similarly to the oracle estimate $\tilde{\theta}_{k^*}$ since it is "stable" in the sense that even if the described selection scheme (12), (13) overshoots k^*, the resulting estimate $\hat{\theta}_{\hat{k}}$ is still close to the oracle $\tilde{\theta}_{k^*}$. In fact the attained quality of estimation during "propagation" is not lost at further

steps:

$$L(W^{k^*}, \tilde{\theta}_{k^*}, \hat{\theta}_{\hat{k}})\, \mathbf{I}\{\hat{k} > k^*\} \leq \mathfrak{z}_{k^*}$$

In other words, $\hat{\theta}_{\hat{k}}$ lies in the confidence set of $\tilde{\theta}_{k^*}$. A combination of the propagation and stability property leads to the "oracle" property:

$$\begin{aligned}
\mathsf{E}_{\theta(\cdot)} \log\left\{1 + \frac{|L(W^{k^*}, \tilde{\theta}_{k^*}, \theta)|^r}{\mathfrak{r}_r}\right\} &\leq \Delta + 1, \\
\mathsf{E}_{\theta(\cdot)} \log\left\{1 + \frac{|L(W^{k^*}, \tilde{\theta}_{k^*}, \hat{\theta}_{\hat{k}})|^r}{\mathfrak{r}_r}\right\} &\leq \Delta + \alpha + \log\left\{1 + \frac{\mathfrak{z}_{k^*}}{\mathfrak{r}_r}\right\},
\end{aligned}$$

for θ with $\Delta(W^k, \theta) \leq \Delta$ and $k \leq k^*$. These bounds show that the risk of estimating adaptively is composed into three parts: the SMB, the false alarm rate, and a small term corresponding to the risk of overshooting.

7.2 Discretization

We now prove the connection of the discrete $AR(3)$ model in (2) and the $CAR(3)$ model in (1) is proved by deriving an analytical link between $X_{k(t)}$ and the lagged deseasonalised temperatures up to time $t-L$. $X_{k(t+L)}$ is approximated by Euler discretization. For example, the step length to be Δ, and observation number to be $N(\Delta)$, for $L = 3\Delta$, let $e_t \stackrel{\text{def}}{=} B_{t+\Delta} - B_t$ and a time step of length one $\Delta_t \stackrel{\text{def}}{=} (t + \Delta) - t = \Delta$, $X_{1(t+3\Delta)}$ is obtained by iteratively

substituting $X_{3(t)}$ from the following discretization:

$$\begin{aligned}
X_{1(t+\Delta)} - X_{1(t)} &= X_{2(t)}\Delta \\
X_{2(t+\Delta)} - X_{2(t)} &= X_{3(t)}\Delta \\
X_{3(t+\Delta)} - X_{3(t)} &= -\alpha_3(\Delta)X_{1(t)}\Delta - \alpha_2(\Delta)X_{2(t)}\Delta - \alpha_1(\Delta)X_{3(t)}\Delta + \sigma_t(\Delta)e_t \\
&\dots \\
X_{1(t+3\Delta)} - X_{1(t+2\Delta)} &= X_{2(t+2\Delta)}\Delta \\
X_{2(t+3\Delta)} - X_{2(t+2\Delta)} &= X_{3(t+2\Delta)}\Delta \\
X_{3(t+3\Delta)} - X_{3(t+2\Delta)} &= -\alpha_3(\Delta)X_{1(t+2\Delta)}\Delta - \alpha_2(\Delta)X_{2(t+2\Delta)}\Delta - \alpha_1(\Delta)X_{3(t+2\Delta)}\Delta \\
&+ \sigma_{t+2\Delta}(\Delta)e_{t+2\Delta}. \qquad (27)
\end{aligned}$$

Rearranging the above equations, we have a $AR(3)$ discrete time model that is linked to (2):

$$\begin{aligned}
X_{1(t+3\Delta)} &= \underbrace{\{3-\alpha_1(\Delta)\}}_{\beta_1(\Delta)} X_{1(t+2\Delta)} + \underbrace{\{2\alpha_1(\Delta) - \alpha_2(\Delta) - 3\}}_{\beta_2(\Delta)} X_{1(t+\Delta)} \\
&+ \underbrace{\{-\alpha_1(\Delta) + \alpha_2(\Delta) - \alpha_3(\Delta) + 1\}}_{\beta_3(\Delta)} X_{1(t)} + \sigma_{t+2\Delta}(\Delta)e_t. \qquad (28)
\end{aligned}$$

Let us define $\mathbf{\Gamma}(\Delta) \stackrel{\text{def}}{=} \{\alpha_1(\Delta), \alpha_2(\Delta), \alpha_3(\Delta), \sigma_t(\Delta)\}^\top$. According to Pedersen (1995), Broze, Scaillet and Zakoian (1998), under some regularity assumption, the estimated parameter in discrete time $\hat{\mathbf{\Gamma}}(\Delta)$ converge to the continuous time parameter $\mathbf{\Gamma}$ as $\Delta \to 0$:

$$\lim_{\Delta \to 0} \lim_{N(\Delta) \to \infty} |\hat{\mathbf{\Gamma}}(\Delta) - \mathbf{\Gamma}| = 0 \qquad \text{a.s.}$$

The continuous time process (1) is Markov and therefore allows standard applications of pricing tools. The last three columns of Table 4 display the $CAR(3)$ parameters $\hat{\alpha}(1)$. Note that in (1) changes in the deseasonalized temperature are regressed against the deseasonalized temperatures itself. This leads to the fact, that the temperature tends toward a seasonal

function, which is not the case in the CD model.

References

Benth, F. E. and Benth, S. (2011). Weather derivatives and stochastic modelling of temperature, *International Journal of Stochastic Analysis* **2011**: 1–21.

Benth, F. E. and Benth, S. (2012). A critical view on temperature modelling for application in weather derivatives markets, *Energy Economics* **34**: 592–602.

Benth, F. E., Benth, S. and Koekebakker, S. (2007). Putting a price on temperature, *Scandinavian Journal of Statistics* **34**: 746–767.

Benth, F. E., Härdle, W. K. and López Cabrera, B. (2011). Pricing asian temperature risk, *in* P. Cizek, W. K. Härdle and R. Weron (eds), *Statistical Tools for Finance and Insurance*, 2 edn, Springer Verlag Heidelberg.

Broze, L., Scaillet, O. and Zakoian, J.-M. (1998). Quasi-indirect inference for diffusion processes, *Econometric Theory* **14**(02): 161–186.

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

Chen, Y., Härdle, W. K. and Pigorsch, U. (2010). Localized realized volatility modelling, *Journal of the American Statistical Association* **105(492)**: 1376–1393.

Cízek, P., Härdle, W. K. and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroschedasticity models, *The Econometrics Journal* **12**: 248–271.

Davis, M. (2001). Pricing weather derivatives by marginal value, *Quantitative Finance* **1**: 305–308.

Granger, C. W. J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* **11**: 399–421.

Härdle, W. K. and López Cabrera, B. (2012). The implied market price of weather risk, *Applied Mathematical Finance* **19(1)**: 59–95.

Horst, U. and Mueller, M. (2007). On the spanning property of risk bonds priced by equilibrium, *Mathematics of Operation Research* **32(4)**: 784–807.

Karatzas, I. and Shreve, S. (2001). *Methods of Mathematical Finance*, Springer Verlag, New York.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models, *The Annals of Statistics* **32(2)**: 577–602.

Pedersen, A. R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* pp. 257–279.

Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation, *Probability Theory and Realated Fields* **135**: 335–362.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value at risk, *The Annals of Statistics* **37**(3): 1405–1436.