# SiZer Analysis for the Comparison of Time Series

Cheolwoo Park[1], Amy Vaughan[1], Jan Hannig[2], Kee-Hoon Kang[3]

**Abstract**

SiZer (SIgnificant ZERo crossing of the derivatives) is a scale-space visualization tool for statistical inferences. In this paper we introduce a graphical device, which is based on SiZer, for the test of the equality of the mean of two time series. The estimation of the quantile in a confidence interval is theoretically justified by advanced distribution theory. The extension of the proposed method to the comparison of more than two time series is also done using residual analysis. A broad numerical study is conducted to demonstrate the sample performance of the proposed tool. In addition, asymptotic properties of SiZer for the comparison of two time series are investigated.

*Key words:* **Comparison of multiple time series, Local linear smoothing, Multiple testing adjustment, Bandwidth, Autocovariance function, Weak convergence.**

## 1 Introduction

SiZer, SIgnificant ZERo crossing of the derivatives, was developed by Chaudhuri and Marron (1999), as an exploratory data analysis tool that is a more advanced version of a basic statistical graphic. SiZer can take data and test it against an assumed model or an estimated model and error structure. As a color coded tool, it can identify not only local extrema, such as peaks and valleys occur, but also where they are precisely located. These extrema are denoted as significant features by SiZer via color changes on either side of the zero crossing of the derivative when it is determined that there are significantly increasing or decreasing portions of the SiZer map. These significant features are denoted to represent real trends in the data and not merely sampling noise artifacts. SiZer speeds up one's ability to determine where features are "really present" in a dataset, and it enables inexperienced statistical analysts to make inferences in gray areas where significance might be questionable.

SiZer takes a nonparametric approach at smoothing curves that can be viewed as a progression of information through $x$ values; $y$ values, which can include time and space; and a number of different bandwidths. It is this vast range of bandwidths that makes SiZer especially unique. In a departure

[1]Department of Statistics, University of Georgia, Athens, GA 30602, USA.

[2]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

[3]Corresponding author. Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Korea. TEL: +82-31-3304486, FAX : +82-31-3304566, E-mail : khkang@hufs.ac.kr

from classical nonparametric curve estimation, which focuses on finding the optimal bandwidth at which to view an image, SiZer takes a scale-space approach and analyzes the data at numerous levels of resolution. Looking at our kernel estimated smooths at these multiple bandwidths allows us to sift out all of the information that is available at different levels of resolution and perform statistical inference at various stages of scrutiny. These multiple bandwidths also move us from the classical approach of finding significant features amidst noisy data of the 'true underlying curve' to finding them in the 'curve at that given level of resolution'.

There are some recent literature on SiZer for nonparametric inference. Hannig and Marron (2006) proposed a new method to reduce spurious pixels in the SiZer map and thus improve inference by replacing the quantile for the confidence interval that was previously based on the idea of independent blocks to a quantile which involved the use of advanced distribution theory. Park and Kang (2008) proposed a SiZer tool which is capable of comparing multiple curves based on their *differences* of smooths when the observed data are independent. Park, Marron, and Rondonotti (2004) proposed a dependent SiZer that does not make the assumption of independent errors. It can detect which features in the SiZer map should not be declared significant because they are in fact attributable to the presence of dependence in the dataset and not a particular trend. This dependent SiZer extends the methodology into time series data and uses an assumed autocovariance function when performing goodness of fit tests. The importance of a goodness of fit test is that it allows us to see how our data differ from the assumed model. But, it is often difficult in non-simulated data to assume the exact autocovariance structure that should be used as the 'true' model. This led Rondonotti, Marron, and Park (2007) to extend the SiZer for time series and develop a method that was more flexible and able to consider the type of dependence structure present in the data in order to detect significant features by not assuming but instead estimating the autocovariance function. Park, Hannig, and Kang (2009) developed an additional improved version of SiZer for time series by incorporating the extreme value theory previously proposed by Hannig and Marron (2006) in order to get a quantile that reduces the number of undesirable spurious pixels. They also proposed a new autocovariance estimator using a differenced time series and it does not rely on pilot bandwidths and residuals from an estimate as in Rondonotti, Marron, and Park (2007). Recently, Bayesian versions of SiZer have been proposed, which include Godtliebsen and Øigård (2005), Erästö and Holmström (2005), and Øigård, Rue and Godtliebsen (2006). The inference is based on finite difference quotients or derivatives, which depends on the selected prior model for the underlying curve.

The problem of testing the equality of nonparametric regression curves with independent errors has been widely studied in the literature. Those include Härdle and Marron (1990), Hall and Hart (1990), Delgado (1993), Kulasekera (1995), Bowman and Young (1996), Kulasekera and Wang (1997), Neumeyer and Dette (2003), Munk and Dette (1998), Dette and Neumeyer (2001), and Pardo-Fernández et al. (2007). Koul and Stute (1998) and Li (2006) studied fitting a regression function in the presence of long memory. To our knowledge the problem of testing equality of two

time series by regression approach has not been considered much.

In order to compare multiple time series, this paper aims to develop a SiZer tool, which is based on regression function estimation. This is an extension of the works regarding the existing SiZer for time series (Rondonotti et al., 2007, and Park et al., 2008) since they are applicable to only one time series. Moreover, this is also an advancement of Park and Kang (2008) since they considered only the independent case. This proposed tool gives insightful information about the differences between the curves by combining statistical inference with visualization. The method presented here not only keeps the advantages of the original SiZer tools, but also extends their usefulness to a broader range of scientific problems.

This paper is organized as follows. Section 2 describes a SiZer for the comparison of two time series. Section 3 describes a SiZer for the comparison of more than two time series. Section 4 investigates the finite sample performances of the proposed method via several simulated examples. Applications to real data are illustrated in Section 5. In Section 6, asymptotic properties of SiZer for the comparison of two time series are investigated. The quantile for constructing confidence intervals in Section 2 is derived in Section 7.

## 2 Comparison of two time series

We start with comparing two time series based on the difference of two kernel estimates. For this reason, in this paper, SiZer stands for SIgnificance of ZERo crossing of the *differences*.

A statistical challenge in this problem is testing whether there is any statistically significant differences between these time series. Suppose that we have $2n$ observations from the following regression models:

$$Y_{ij} = f_i(j) + \sigma_i \varepsilon_{ij}, \quad j = 1, \ldots, n, \ i = 1, 2, \tag{2.1}$$

where the $\varepsilon_{ij}$'s are dependent random errors with mean 0, variance 1, and $Cov(\varepsilon_{ij}, \varepsilon_{ik}) = \gamma_i(|j-k|)$ for all $i = 1, 2$, $j, k = 1, \ldots, n$, $f_i$ is the unknown regression function of the $i$th sample and $\sigma_i^2$ is the variance function of the $i$th sample ($i = 1, 2$). We assume that $\varepsilon_{1j}$ and $\varepsilon_{2j}$ are independent of each other.

Our main concern is to develop a graphical device for testing the following hypothesis of the equality of mean regression functions

$$H_0 : f_1 = f_2$$

when the errors are weakly correlated.

SiZer takes an approach that applies the local linear fitting method, see e.g. Fan and Gijbels (1996), for obtaining a family of kernel estimates in a regression setting. Precisely, at a particular point $x_0$, $\hat{f}_{i,h}(x_0)$ ($i = 1, 2$) are obtained by fitting lines

$$\beta_{i0} + \beta_{i1}(x_0 - j)$$

to the $(j, Y_{ij})$, with kernel weighted least squares. Then, $\hat{f}_{i,h}(x_0) = \hat{\beta}_{i0}$ $(i = 1, 2)$ where $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1})'$ minimizes

$$\sum_{j=1}^{n} \{Y_{ij} - (\beta_{i0} + \beta_{i1}(x_0 - j))\}^2 K_h(x_0 - j), \tag{2.2}$$

where $K_h(\cdot) = K(\cdot/h)/h$. $K$ is a kernel function, usually a symmetric probability density function. In this paper, we use a Gaussian kernel. Since the solution of (2.2) provides estimates of a regression function for different bandwidths, we can construct the family of smooths parameterized by $h$ and the confidence intervals of the difference of two time series.

In SiZer analysis, the hypotheses we are testing are

$$H_0 \ : \ f_{1,h}(x_0) = f_{2,h}(x_0) \quad \text{vs.} \quad H_1 \ : \ f_{1,h}(x_0) \neq f_{2,h}(x_0) \tag{2.3}$$

for a fixed time point $x_0$. Here, $f_{i,h}(x) \equiv E\hat{f}_{i,h}(x)$ is the scale-space version of $f_i(x)$, $i = 1, 2$.

SiZer visually displays the significance of differences between two regression functions in families of smooths $\{\hat{f}_{i,h}(x), i = 1, 2\}$ over both location $x$ and scale $h$, using a color map. It is based on confidence intervals for $\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)$, which will be defined soon, and uses multiple comparison level adjustment. Each pixel shows a color that gives the result of a hypothesis test in (2.3) at the point indexed by the horizontal location $x$, and by the bandwidth corresponding to the row $h$. At each $(x, h)$, if the confidence interval is above (below) 0, meaning that the curves are significantly different, i.e., $f_{1,h}(x) > f_{2,h}(x)$ $(f_{1,h}(x) < f_{2,h}(x))$, then that particular map location is colored blue (red, respectively). On the other hand, if the confidence interval contains 0, meaning that the curves are not significantly different, then that map location is given purple. Finally, if there are not enough data points to carry out the test, then no decision can be made and the location is colored gray. To determine the gray areas, as in Chaudhuri and Marron (1999), we define the estimated effective sample size (ESS), for each $(x, h)$ as

$$\text{ESS}(x, h) = \frac{\sum_{j=1}^{n} K_h(x - j)}{K_h(0)}.$$

If $\text{ESS}(x, h) < 5$, then the corresponding pixel is colored gray.

Confidence intervals for $f_{1,h}(x) - f_{2,h}(x)$ are of the form

$$\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x) \pm q \cdot \widehat{SD}(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)), \tag{2.4}$$

where $q$ is an appropriate quantile, and the standard deviation is estimated as discussed soon. For the approximation of the quantile, Chaudhuri and Marron (1999) suggested several methods including pointwise Gaussian quantiles, number of independent blocks, and bootstrap. Recently, Hannig and Marron (2006) improved the multiple comparison tests using advanced distribution theory. A similar calculation can be done for the comparison of two time series. As a result, the quantile for significance level $\alpha$ is defined as

$$q = \Phi^{-1}\left(\left(1 - \frac{\alpha}{2}\right)^{1/(\theta g)}\right)$$

where $\Phi$ is the standard normal distribution function and $g$ is the number of bins. The "cluster index" $\theta$ is given by

$$\theta = 2\Phi\left(\sqrt{I \log g}\,\frac{\tilde{\Delta}}{h}\right) - 1,$$

where

$$I = \frac{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta))\, e^{-s^2/4} \frac{2-s^2}{8}\, ds}{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta))\, e^{-s^2/4}\, ds}.$$

Here, $\tilde{\Delta}$ denotes the distance between the pixels of the SiZer map, and $\gamma_1$ and $\gamma_2$ are the autocovariance functions of the first and the second time series, respectively. We use this quantile in our implementation and the brief derivation of the cluster index $\theta$ is provided in Section 7.

For the estimation of the standard deviation, note that $\hat{f}_{i,h}(t)$ obtained from (2.2) can be written as

$$\hat{f}_{i,h}(x) = \frac{1}{n}\sum_{j=1}^{n} w_n(h, x, j) Y_{ij}$$

where

$$w_n(h, x, j) = \frac{\{\hat{s}_2(x; h) - \hat{s}_1(x; h)(x - j)\} K_h(x - j)}{\hat{s}_2(x; h)\hat{s}_0(x; h) - \hat{s}_1(x; h)^2}$$

and

$$\hat{s}_r(x; h) = \frac{1}{n}\sum_{j=1}^{n} (x - j)^r K_h(x - j).$$

Then, by independence

$$Var(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)) = Var(\hat{f}_{1,h}(x)) + Var(\hat{f}_{2,h}(x)),$$

and

$$Var(\hat{f}_{i,h}(x)) = \frac{\sigma_i^2}{n^2}\sum_{j=1}^{n} (w_n(h, x, j))^2 + \frac{2}{n^2}\sum_{j<k} w_n(h, x, j) w_n(h, x, k) \gamma_i(k - j). \tag{2.5}$$

In order to construct the confidence interval in (2.4) we need to estimate the autocovariance function $\gamma_i$ in (2.5). Park, Hannig, and Kang (2009) proposed a new estimator of this $\gamma_i$ using a differenced time series in one sample case. Let $\mathbf{e}_i$ be the $i$th differenced time series ($i = 1, 2$), i.e., $\mathbf{e}_i = A\mathbf{y}_i$ where $A = (a_{jk})$ is the difference matrix, e.g.,

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

if the first difference is used. A simple calculation shows for all $j, k$

$$\begin{aligned}\mathrm{Cov}(e_{ij}, e_{ik}) &= \sum_{l=1}^{n} a_{j,l}a_{k,l}\gamma_i(0) + \sum_{l=1}^{n-1}(a_{j,l}a_{k,l+1} + a_{j,l+1}a_{k,l})\gamma_i(1) \\ &\quad + \cdots + (a_{j,1}a_{k,n} + a_{j,n}a_{k,1})\gamma_i(n-1).\end{aligned}$$

From this we can set a regression setting

$$\begin{aligned}e_{ij}e_{ik} &= \sum_{l=1}^{n} a_{j,l}a_{k,l}\gamma_i(0) + \sum_{l=1}^{n-1}(a_{j,l}a_{k,l+1} + a_{j,l+1}a_{k,l})\gamma_i(1) \\ &\quad + \cdots + (a_{j,1}a_{k,n} + a_{j,n}a_{k,1})\gamma_i(n-1) + \delta_{jk}.\end{aligned}$$

We assume that the regression function was smooth enough so that $E(\delta_{jk}) \approx 0$. Estimating $\gamma_i$ by the least squares method, i.e. by minimizing

$$\begin{aligned}\sum_{j,k}\Bigg(e_{ij}e_{ik} - \sum_{l=1}^{n} a_{j,l}a_{k,l}\gamma_i(0) - \sum_{l=1}^{n-1}(a_{j,l}a_{k,l+1} + a_{j,l+1}a_{k,l})\gamma_i(1) \\ - \cdots - (a_{j,1}a_{k,n} + a_{j,n}a_{k,1})\gamma_i(n-1)\Bigg)^2\end{aligned} \tag{2.6}$$

fails because the least squares problem in (2.6) does not lead to a unique solution.

Park, Hannig, and Kang (2009) proposed to regularize the problem (2.6). First, since $\gamma_i(0) \geq |\gamma_i(j)|$ for each $i$ and $j$, we consider only such solutions. Additionally, we regularize the least squares problem by introducing the penalty $\lambda\sum_{l=1}^{n-1} l\gamma_i(l)^2$. The weight $l$ is motivated by the belief that the covariance $\gamma_i(l)$ should be decaying as $l$ increases.

This leads to the following constrained ridge regression

$$\begin{aligned}\underset{\gamma_i \in R_i}{\arg\min}\Bigg\{\sum_{j,k}\Bigg(e_{ij}e_{ik} - \sum_{l=1}^{n} a_{j,l}a_{k,l}\gamma_i(0) - \sum_{l=1}^{n-1}(a_{j,l}a_{k,l+1} + a_{j,l+1}a_{k,l})\gamma_i(1) \\ - \cdots - (a_{j,1}a_{k,n} + a_{j,n}a_{k,1})\gamma_i(n-1)\Bigg)^2 + \lambda\sum_{l=1}^{n-1} l\gamma_i(l)^2\Bigg\},\end{aligned}$$

where $R_i = \{\gamma_i \,:\, \gamma_i(0) \geq |\gamma_i(j)|, j = 1, \ldots, n-1\}$. We have investigated several choices of $\lambda$ and found that $\lambda = 1$ works well as long as the time series is weakly to moderately dependent.

## 3 SiZer for the comparison of multiple time series

This section is devoted to testing the equality of $k(>2)$ time series. The model (2.1) becomes

$$Y_{ij} = f_i(j) + \sigma_i\varepsilon_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots, n.$$

We consider testing the following scale-space version of the hypotheses

$$H_0 \,:\, f_{1,h}(x_0) = f_{2,h}(x_0) = \cdots = f_{k,h}(x_0) \quad \text{vs.} \quad H_1 \,:\, \textit{not} \;\; H_0. \tag{3.1}$$

However, the extension of the approach in Section 2 is not a straightforward application for this testing problem. Therefore, we propose to compare two sets of residual time series under the null and alternative hypotheses, respectively. To accomplish this, first we obtain two residual sets by fitting local linear estimates under the null and alternative hypotheses in (3.1), and then compare their time series by the tool developed in Section 2. One set of residuals will reflect the difference between each individual dataset and its estimated function at a specific bandwidth. The second set of residuals will compare the individual datasets to a common overall estimated function by pooling $k$ sets of time series. Under the null hypothesis, the residuals obtained by fitting an individual mean function for each data set should behavior similarly to the residuals computed by fitting a common estimator function. In this way, we convert the comparison of multiple time series into the comparison of two time series. A similar idea was used in Park and Kang (2008) for the independent case.

For obtaining the residuals, one could use a pilot bandwidth $h_p$, which is different from the bandwidth $h$ used in constructing a SiZer map. For simplicity, however, we take $h_p = h$ in our analysis.

Let $\hat{f}_h(\cdot)$ be the local linear estimator of the common scale-spaced regression function $f_h(\cdot)$ under $H_0$, which has the following form:

$$\hat{f}_h(x) = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} w_n(h, x, j) Y_{ij},$$

where $w_n(h, x, j)$'s are the local linear weights. Let $Y_{ij} - \hat{f}_{i,h}(j)$ be the estimate of the error $\varepsilon_{ij}$ from the $i$th population and let $Y_{ij} - \hat{f}_h(j)$ be the estimate of the same quantity under the null hypothesis in (3.1). The idea is that if $H_0$ is true, $Y_{ij} - \hat{f}_{i,h}(j)$ and $Y_{ij} - \hat{f}_h(j)$ would be quite similar time series, thus their difference in a SiZer map would reflect the color purple. Hence, we can check the equality of $k$ time series by comparing these two sets of residual time series with the tool proposed in Section 2. We will take a closer look at this method's performance in simulations and real data analyses that will soon be presented.

# 4 Simulation

The first part of this section illustrates the simulated examples when the dependence structure is previously known and the second part deals cases when the dependence structure is unknown for two time series. A simulation study for three time series is done in the third part.

## 4.1 Two time series: when the dependence structure is known

This section shows the performance of SiZer when the autocovariance function is given in advance. This is an extension of Park, Marron, and Rondonotti (2004) for two time series, and it is particularly useful when the dependence structure is known from previous studies.

In our simulation we consider various combinations of error structures and mean regression functions, but we report only a few of them to save space. Three simulated examples are provided and each example has the sample size $n = 100$. The first example has the same constant mean 0:

$$\text{(i)} \qquad Y_{ij} = \varepsilon_{ij}, \quad j = 1, \ldots, n, \ i = 1, 2.$$

For the second example, one time series has a sine curve as regression function and the other has mean 0:

$$\text{(ii)} \qquad Y_{1j} = 4\sin(6\pi j/n) + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}.$$

The third example studies two different regression functions and has the following regression models:

$$\text{(iii)} \qquad Y_{1j} = 4\sin(6\pi j/n) + 3j/n + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = 4\sin(6\pi j/n) + \varepsilon_{2j}.$$

We consider two combinations of error structures, MA(1) versus AR(1), and MA(1) versus MA(5), for weakly correlated data and strongly correlated one, respectively.. The correct SiZer plots would show no significant difference for the first example, a sine trend for the second, and a linear trend for the third.

Figure 1 about here.

Figure 1 displays SiZer plots with MA(1) and AR(1) for the three examples. In the top two panels, the green dots are actual data points and the thin blue curves display the family of smooths, i.e. $\hat{f}_{i,h}(x)$ for $i = 1, 2$. The SiZer maps in the third panels report the equality test of the two time series by investigating the confidence intervals in (2.4) at each $(x, h)$. The horizontal locations in the SiZer map are the same $x$ values as in the top panels, and the vertical locations in the SiZer plot correspond to the logarithm of bandwidths of the family of smooths shown as thin blue curves in the top panels. Each pixel in the SiZer map shows a color that gives the result of a hypothesis test for the significance of the differences between the thin blue curves in family plot 1 minus family plot 2, at the point indexed by the horizontal location, and at the bandwidth corresponding to that row. The SiZer map in Figure 1 (a) shows only purple, meaning no significant difference, as expected. We see also that when the two regression curves are different, the SiZer maps correctly capture the differences. The SiZer map in Figure 1 (b) shows positive (blue) and negative (red) differences along the sine curve, but also has some spurious pixels present. The map in Figure 1 (c) flags a rough linear trend as significant. From these three simulations, we show that SiZer for the comparison of two time series performs well in its ability to detect significant trends through assumed dependence structure for weakly correlated data.

Figure 2 about here.

Figure 2 displays SiZer plots with MA(1) and MA(5) for the three examples. The results are similar to Figure 1, with (a) being correctly marked completely purple and (b) has the sine trend

marked even more cleanly here with no spurious pixels. Figure 2 (c) once again detects a general linear trend, with a couple of spurious pixels near the center of the picture. Thus, SiZer performs reasonably well for cases in which either of the time series has a weak or a stronger correlation, but it does need some improvement in clearly delineating the linear trend as the significant signal.

### 4.2 Two time series: when the dependence structure is unknown

In real world, we can hardly assume the true autocovariance structure in advance. Therefore, we examine the performance of our approach with the estimated autocovariance functions in Section 2 by repeating the simulation study in Section 4.1. Comparing the results with those in Section 4.1 enables us to assess the performance of our autocovariance function estimator.

Figure 3 about here.

Figure 3 displays SiZer plots with MA(1) and AR(1) for the three examples. As one can see from the plots, similar to Figure 1, SiZer flags no trend for Figure 3 (a) and thus the entire map is purple. In Figure 3 (b) the SiZer map also catches all of the important trends given in a sine curve although it is not quite as clearly evenly partitioned as is desired, with some spurious pixels in the upper left hand corner. In Figure 3 (c), it is very similar to Figure 1 which has the true autocovariance function, and it does detect a strong difference in the upper portion of the map that is attributed to the linear trend.

Figure 4 about here.

Figure 4 displays SiZer plots with MA(1) and MA(5) for the same three trend examples. Here again we can make almost the same conclusions as those plots in Figure 2. In Figure 4 (a), it is correctly colored purple for the presence of no trend and in Figure 4 (b), we can see that the SiZer map captures all of the changes in the sine trend and our only misdiagnoses are in the top right hand corner where we can see some spurious pixels. Figure 4 (c) also shows again that the positive linear trend from the first plot is identified by the SiZer map. Here with these two figures we have seen that SiZer succeeds in capturing the important differences in two correlated time series while estimating the autocovariance function. In addition, it also does a very fair job of highlighting the differences in trend whether both time series have weak correlation or if one of the two time series has a stronger correlation.

### 4.3 Multiple time series

In this section, we simulate two examples to compare three different time series. In each example, the three time series are generated from $N(0, 1)$, MA(1), and AR(1), respectively, with the length $n = 100$. In the first example, the mean regression functions are all zero. So, each graph should demonstrate that there is no signal and if the time series can all accurately be accounted for, the

behavior of residuals under each of the three estimation functions should be the same as that of the commonly estimated function. These two mean functions of residuals should therefore have leave a difference of nothing and the lack of trend would leave us with three SiZer maps of no significant trend.

Figure 5 about here.

In Figure 5, we see that the first column of graphs shows the generated time series and their family of smooths plots. The second column shows the SiZer maps comparing two sets of residual time series. In other words, for $i = 1, 2, 3$, the $i$th row of the second column corresponds to the SiZer map comparing $Y_{ij} - \hat{f}_{i,h}(j)$ and $Y_{ij} - \hat{f}_h(j)$. All purple colors indicate that the mean functions of the residual time series are indeed similar according to SiZer analysis.

In the second example, the error structures remain the same as in the first, but we add the sine curve $f_1(x) = \sin(6\pi x)$ to the first sample. The last two samples remain with no signal, only the correlated error structure.

Figure 6 about here.

We can see that Figure 6 shows some significant features in the SiZer maps, which implies the differences of the mean functions. The first one shows more significant trends since it is different from the others, and the trend clearly suggests the presence of the sine curve that was inserted into the first sample.

## 5 Real data analysis

This section is devoted to illustrating our procedure applied to real data.

Figure 7 about here.

*Example 1.* This first example involves the yields of the 3-month, 6-month, and 12-month Treasury bills. The data set was taken from July 1959 to August 2001. In order to decrease the sample size, we have take the average of every 2 consecutive months and used that as our data, causing no change in trend. The original data can be seen as examples in sources such as Fan and Yao (2003). We can see the almost identical structure in the family plots of all 3 time periods. In Figures 7 (a), (b), and (c) we see that the almost identical correlation of the rates is detected by the SiZer maps, all of which are purple, indicating no significant difference between any pair of the time periods.

Figure 8 about here.

*Example 2.* This example displays the long-term rates for US, Canada and Japan from January

1980 to December 2000. Before plotting, the global mean has been taken out for each country so that all of the data is centered in order to compare the relative trends of interest rates between the countries. A few comments will be made relative to the article by Christiansen and Pigott (1997) to point out the similarities in their assessed trends to the ones detected by SiZer. We see in Figure 8 (a) that the long-term interest rates for the US and Canada moved quite closely together from approximately 1993-1995, despite different business cycle positions at those times. This is indicated by the purple marking of no significance between the red and final blue highlighted portions. We also can confirm in the SiZer map the events of the fall of the Canadian rates to just below the US rates for the first time in over a decade around 1996 indicated by the final blue difference section in the map.

In Figure 8 (b), we can see that in the period from 1982 to mid 1984 the US rates rose as the Japanese rates were falling, believed by the authors to be caused in part by the effects of US fiscal expansion in raising the demand for domestic savings relative to its supply. This is indicated by the blue highlighted pixels to the left of the graph indicating this early 1980's time period. In Figures 8 (a) and (b) we can see significant divergences in the interest rates in the late 1980's as US rates begin to fall back, rates in Canada and Japan are increasing. In both plots, the larger values of Canada and Japan cause a significant negative difference, denoted red in the middle of both plots. We can see this short-term similarity between Canada and Japan in Figure 8 (c), however the graph is clearly dominated by the more rapid descent of the Canadian rates through the overall decrease of both countries.

Figure 9 about here.

It would also be interesting to compare three yields of three countries at the same time, as in Example 1 and Example 2, respectively. To save space we only report the result of Example 2 for multiple comparison. We can see that in Figure 9, there are differences that occur within each SiZer map, denoting that there are significant differences present. We have seen in Figure 8 that there existed pairwise differences between all of the countries. The presence of these differences are also correctly detected when we compare each set of residuals from each country's individual estimated function to the residuals from the overall estimation.

*Remark.* Another approach using ANOVA type statistics can be developed for the comparison of multiple curves. The test statistics for comparing the curves at $x$ can be roughly written as

$$(Constant) \times \frac{\sum_{i=1}^{k}(\hat{f}_{i,h}(x) - \hat{f}_h(x))^2}{\sum_{i=1}^{k}\sum_{j=1}^{n}(Y_{ij} - \hat{f}_{i,h}(j))^2 K_h(x-j)}$$

where $\hat{f}_{i,h}$ is a local linear fit using $i$th sample and $\hat{f}_h$ using the combined sample under the null hypothesis. This statistic mimics the ratio of variations from the model and the error in ANOVA. To conduct a test, one needs to find the approximate distribution of this statistic and its degrees

of freedom. Also, an appropriate multiple adjustment needs to be designed for SiZer. If some differences are found among the curves, multiple pairwise comparisons can be performed as done in ANOVA analysis. We propose this approach as our future work.

## 6 Asymptotic results

In this section we study statistical convergence of the difference between the empirical and the theoretical scale space surfaces, which provides theoretical justification of SiZer for the comparison of two time series in scale space. Chaudhuri and Marron (2000) addressed this issue based on one independent sample and Park, Hannig, and Kang (2009) extended it to single correlated data. Here, we extend it to the case of comparing two correlated samples.

The first theorem provides the weak convergence of the empirical scale space surfaces and their differences to their theoretical counterpart. The second theorem states the behavior of the difference between the empirical and the theoretical scale space surfaces under the supremum norm and the uniform convergence of the empirical version to the theoretical one.

Let $I$ and $H$ be compact subintervals of [0,$\infty$) and (0,$\infty$), respectively. Let

$$\hat{g}_h(x) = \frac{1}{n}\sum_{j=1}^{n} Z_j w_n(h, x, j)$$

where $Z_j = Y_{1j} - Y_{2j}$. The following set of assumptions are needed for the following theorems.

**(A.1)** The errors $(\varepsilon_{i1}, \varepsilon_{i2}, ...)$ in (2.1) are stationary, $\phi$-mixing with the mixing function $\phi(j)$ satisfying $\sum_{j=1}^{\infty} \phi(j)^{1/2} < \infty$. (See for example Doukhan (1994) for definition of $\phi$-mixing.)

**(A.2)** The errors have a bounded moment $E\{|\varepsilon_{ij}|^{2+\rho}\} < \infty$ for some $\rho > 0$.

**(A.3)** For integer $n \geq 0$, as $n \to \infty$

$$\frac{1}{n}\left[\sum_{j=1}^{n}\sum_{k=1}^{n}(\gamma_1(|j-k|) + \gamma_2(|j-k|))w_n(h_1, x_1, j)w_n(h_2, x_2, k)\right]$$

converges to a covariance function $cov(h_1, x_1, h_2, x_2)$ for all $(h_1, x_1)$ and $(h_2, x_2) \in H \times I$.

**(A.4)** $n^{-(1+\rho/2)}\{\max_{1\leq j\leq n} |w_n(h, x, j)|^{\rho}\} \sum_{j=1}^{n} w_n(h, x, j)\}^2 \to 0$ for all $(h, x) \in H \times I$.

**(A.5)** $w_n(h, x, j)w_n(h, x, k)$ will be uniformly dominated by a positive finite number $M$.

**(A.6)**

$$\left\{\frac{\partial w_n(h, x, j)}{\partial x}\right\}\left\{\frac{\partial w_n(h, x, k)}{\partial x}\right\}, \left\{\frac{\partial w_n(h, x, j)}{\partial h}\right\}\left\{\frac{\partial w_n(h, x, k)}{\partial h}\right\}$$

and

$$\left\{\frac{\partial w_n(h, x, j)}{\partial x}\right\}\left\{\frac{\partial w_n(h, x, k)}{\partial h}\right\}$$

will be uniformly dominated by a positive finite number $M^*$.

**Theorem 1** *Suppose that assumptions (A.1)-(A.5) are satisfied. Define*

$$U_n(h,x) = n^{1/2}[\hat{g}_h(x) - E\{\hat{g}_h(x)\}], \quad (h,x) \in H \times I.$$

*As $n \to \infty$, $U_n(h,x)$ converges to a Gaussian process on $H \times I$ with zero mean and covariance function $cov(h_1,x_1,h_2,x_2)$.*

**Proof.** It is enough to show that all the finite dimensional distribution of the process converges weakly to the normal distribution and the process satisfies the tightness condition.

Fix $(h_1,x_1),(h_2,x_2),...,(h_l,x_l) \in H \times I$ and $(t_1,...,t_l) \in (-\infty,\infty)$. Define

$$\begin{aligned} Q_n &= n^{1/2}\sum_{j=1}^{l} t_j[\hat{g}_{h_j}(x_j) - E\{\hat{g}_{h_j}(x_j)\}] \\ &= n^{-1/2}\sum_{p=1}^{n}(\varepsilon_{1p} - \varepsilon_{2p})\sum_{j=1}^{l} t_j w_n(h_j,x_j,p). \end{aligned}$$

Then $E(Q_n)$=0 and

$$\begin{aligned} VAR(Q_n) &= \frac{1}{n}\sum_{j=1}^{l}\sum_{k=1}^{l} t_j t_k \left[\sum_{p=1}^{n}\sum_{q=1}^{n}(\gamma_1(|p-q|) + \gamma_2(|p-q|))w_n(h_j,x_j,p)w_n(h_k,x_k,q)\right] \\ &\to \sum_{j=1}^{l}\sum_{k=1}^{l} t_j t_k cov(h_j,x_j,h_k,x_k) \end{aligned} \tag{6.1}$$

as $n \to \infty$ by assumption (A.3).

Assumptions (A.2) and (A.4) imply that Lyapunov's and hence Lindeberg's condition holds for the terms in $Q_n$. This and assumption (A.1) verify the conditions of the main theorem in Utev (1990) allowing us to conclude that $Q_n$ converges in distribution to a normal random variable with variance given by (6.1). By Cramer-Wold device, the limiting distribution of $U_n(h_j,x_j)$ $(j=1,...,l)$ is the multivariate normal distribution with zero mean and $cov(h_j,x_j,h_k,x_k)$ as the $(j,k)$th entry of the limiting variance-covariance matrix.

We now proceed to the issue of tightness. Fix $h_1 < h_2$ in $H$ and $x_1 < x_2$ in $I$. Then, by Bickel and Wichura (1971) the second moment of increment of $U_n$ is defined by

$$\begin{aligned} &E\{U_n(h_2,x_2) - U_n(h_2,x_1) - U_n(h_1,x_2) + U_n(h_1,x_1)\}^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(\gamma_1(|k-j|) + \gamma_2(|k-j|))D_j D_k \end{aligned} \tag{6.2}$$

where

$$D_j = w_n(h_2,x_2,j) - w_n(h_2,x_1,j) - w_n(h_1,x_2,j) + w_n(h_1,x_1,j).$$

Then, by the assumption (A.5), (6.2) is bounded by

$$C_1(x_2-x_1)^2(h_2-h_1)^2\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(\gamma_1(|j-k|) + \gamma_2(|j-k|)),$$

which is again bounded by $C_2(x_2 - x_1)^2(h_2 - h_1)^2$, since conditions (A.1) and (A.2) imply that $\sup_n n^{-1} \sum_{j=1}^n \sum_{k=1}^n (\gamma_1(|j-k|) + \gamma_2(|j-k|)) < \infty$, c.f., Doukhan (1994), page 45. Then the tightness property of the sequence of processes

$$n^{1/2}[\hat{g}_h(x) - E\{\hat{g}_h(x)\}]$$

on $H \times I$ is implied by the Theorem 3 in Bickel and Wichura (1971). Together with the finite dimensional convergence property, this implies that the theorem holds.

**Theorem 2** *Suppose that assumptions (A.1)-(A.6) are satisfied. As* $n \to \infty$

$$\sup_{x \in I, h \in H} n^{1/2} |\hat{g}_h(x) - E\{\hat{g}_h(x)\}|$$

*converges weakly to a random variable that has the same distribution as that of* $\sup_{x \in I, h \in H} |G(h,x)|$, *where* $G(h,x)$ *is a Gaussian process with zero mean and covariance function* $cov(h_1, x_1, h_2, x_2)$ *so that*

$$P\{G(h,x) \text{ is continuous for all } (h,x) \in H \times I\} = 1,$$

*and consequently* $P\{\sup_{x \in I, h \in H} |G(h,x)| < \infty\}$=1.

**Proof.** Let us denote $D_j^*$ by

$$D_j^* = w_n(h_2, x_2, j) - w_n(h_1, x_1, j).$$

Then,

$$\begin{aligned} E\{U_n(h_2,x_2) - U_n(h_1,x_1)\}^2 &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n (\gamma_1(|j-k|) + \gamma_2(|j-k|)) D_j^* D_k^* \\ &\leq C_3\{(h_2 - h_1)^2 + (x_2 - x_1)^2\}. \end{aligned}$$

Then the rest of the proof can be done the same way in Chaudhuri and Marron (2000) by defining the pseudo metric $d$ by $d\{(h_2,x_2),(h_1,x_1)\} = [E\{G(h_2,x_2) - G(h_1,x_1)\}^2]^{1/2}$.

# 7 Appendix

SiZer uses the local linear smoother defined by (2.2). To color the pixels SiZer checks whether the difference of the estimates of the two regression functions

$$\begin{aligned} \hat{\beta}_{i0} &= c_i^{-1} \left[ \sum_{j=1}^n K_h(x-j) Y_{ij} \right] \left[ \sum_{j=1}^n (x-j)^2 K_h(x-j) \right] \\ &- c_i^{-1} \left[ \sum_{j=1}^n (x-j) K_h(x-j) \right] \left[ \sum_{j=1}^n (x-j) K_h(x-j) Y_{ij} \right], \qquad (7.1) \\ c_i &= \left[ \sum_{j=1}^n K_h(x-j) \right] \left[ \sum_{j=1}^n (x-j)^2 K_h(x-j) \right] - \left[ \sum_{j=1}^n (x-j) K_h(x-j) \right]^2, \end{aligned}$$

for $i = 1, 2$, is significantly different from 0.

$$T_k \approx \sum_{q=1}^{n} w^h_{kp-q}(Y_{1,q} - Y_{2,q}).$$

The form of the $w^h_{kp-q}$ is given in the first term of (7.1). Note that $w^h_{kp-q}$ is proportional to $K_{h/\Delta}(kp - q)$ and thus the weights $w^h_q$ are proportional to the Gaussian kernel with standard deviation $h/\Delta$.

Let $\gamma_1$ be the autocovariance function of the first time series and $\gamma_2$ be the autocovariance function of the second. The full joint distribution of $T_1, ..., T_g$ also depends on the correlation between them. This correlation is approximated by

$$\begin{aligned}
\rho_{j-i} =& corr(T_i, T_j) \\
=& \frac{\sum_q \sum_r w^h_{ip-q} w^h_{jp-r} (\gamma_1(q-r) + \gamma_2(q-r))}{\sum_q \sum_r w^h_q w^h_r (\gamma_1(q-r) + \gamma_2(q-r))} \\
\approx& \frac{\iint K_{h/\Delta}(ip - x)\, K_{h/\Delta}(jp - y)\, (\gamma_1(x-y) + \gamma_2(x-y))\, dxdy}{\iint K_{h/\Delta}(x)\, K_{h/\Delta}(y)(\gamma_1(x-y) + \gamma_2(x-y))\, dxdy} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s)) \int K_{h/\Delta}(ip - s - y)\, K_{h/\Delta}(jp - y)\, dy\, ds}{\int (\gamma_1(s) + \gamma_2(s)) \int K_{h/\Delta}(s + y)\, K_{h/\Delta}(y)\, dy\, ds} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-(ip-jp-s)^2\Delta^2/(4h^2)}\, ds}{\int (\gamma_1(s) + \gamma_2(s)) e^{-s^2\Delta^2/(4h^2)}\, ds} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-[(i-j)\tilde{\Delta} - s\Delta]^2/(4h^2)}\, ds}{\int (\gamma_1(s) + \gamma_2(s)) e^{-s^2\Delta^2/(4h^2)}\, ds}
\end{aligned}$$

We then calculate that

$$\rho_{j,g} = \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-(Cj/\sqrt{\log g} - s)^2/4}\, ds}{\int (\gamma_1(s) + \gamma_2(s))\, e^{-s^2/4}\, ds}.$$

Finally, since $\gamma_i(s)$ is an even function, we get by expanding into a series that

$$\lim_{g\to\infty} \log g(1 - \rho_{k,g}) = k^2 \frac{C^2 \int (\gamma_1(s) + \gamma_2(s))\, e^{-s^2/4} \frac{2-s^2}{8}\, ds}{\int (\gamma_1(s) + \gamma_2(s))\, e^{-s^2/4}\, ds}.$$

Therefore just as in Hannig and Marron (2006) we conclude that in the case of SiZer

$$P\left[\max_{i=1,...,g} T_i \le x\right] \approx \Phi(x)^{\theta g},$$

where the cluster index

$$\theta = 2\Phi\left(\sqrt{I \log g}\, \frac{\tilde{\Delta}}{h}\right) - 1$$

and

$$I = \frac{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta))\, e^{-s^2/4} \frac{2-s^2}{8}\, ds}{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta))\, e^{-s^2/4}\, ds}.$$

## Acknowledgement

The first author was supported by National Security Agency under Grant Number H98230-08-1-0056. The third author was supported in part by the National Science Foundation under Grants No. 0504737 and 0707037. The fourth author was supported by Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-521-C00050).

## References

[1] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, **42**, 1656–1670.

[2] Bowman, A. and Young, S. (1996). Graphical Comparison of Nonparametric Curves. *Applied Statistics*, **45**, 83–98.

[3] Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94**, 807–823.

[4] Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *The Annals of Statistics*, **28**, 408–428.

[5] Christiansen, H. and Pigott, C. (1997). Long-term interest rates in globalised markets. *Economic Department Working Papers*, **175**, 1–44.

[6] Delgado, M. A. (1993). Testing the equality of nonparametric regression curve. *Statistics & Probability Letters*, **17**, 199–204

[7] Dette, H. and Neumeyer, N. (2001). Noparametric analysis of covariance. *The Annals of Statistics*, **29**, 1361–1400.

[8] Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics **85**. Springer, New York.

[9] Erästö, P. and Holmström, L. (2005). Bayesian Multiscale Smoothing for Making Inferences about Features in Scatterplots. *Journal of Computational and Graphical Statististics* **14**, 569–589.

[10] Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

[11] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer.

[12] Godtliebsen, F. and Øigård, T. A. (2005). A visual display device for significant features in complicated signals. *Computational Statistics and Data Analysis* **48**, 317–343.

[13] Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regreeeion. *Journal of the American Statistical Association*, **85**, 1039–1049.

[14] Hannig, J. and Marron, J. S. (2006). Advanced Distribution Theory for SiZer. *Journal of the American Statistical Association*, **101**, 484–499 .

[15] Härdle W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *The Annals of Statistics*, **13**, 63–89.

[16] Koul, H. L. and Stute, W. (1998). Regression model fitting with long memory errors. *Journal of Statistical Planning and Inferences*, **71**, 35–56.

[17] Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, **90**, 1085–1093.

[18] Kulasekera, K. B. and Wang, J. (1995). Smoothing parameter selection for power optimality in testing of regression curves. *Journal of the American Statistical Association*, **92**, 500–511.

[19] Li, F. (2006). Testing for the equality of two nonparametric regression curves with long memory errors. *Communications in Statistics-Simulation and Computation*, **35**, 621–643.

[20] Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *The Annals of Statistics*, **26**, 2339-2368.

[21] Neumeyer, N. and Dette, H. (2003). Noparametric comparison of regression curves: an empirical process approach. *The Annals of Statistics*, **31**, 880–920.

[22] Øigård, T. A., Rue, H. and Godtliebsen, F. (2006). Bayesian multiscale analysis for time series data. *Computational Statistics and Data Analysis* **51**, 1719–1730.

[23] Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, **17**, 1115–1137.

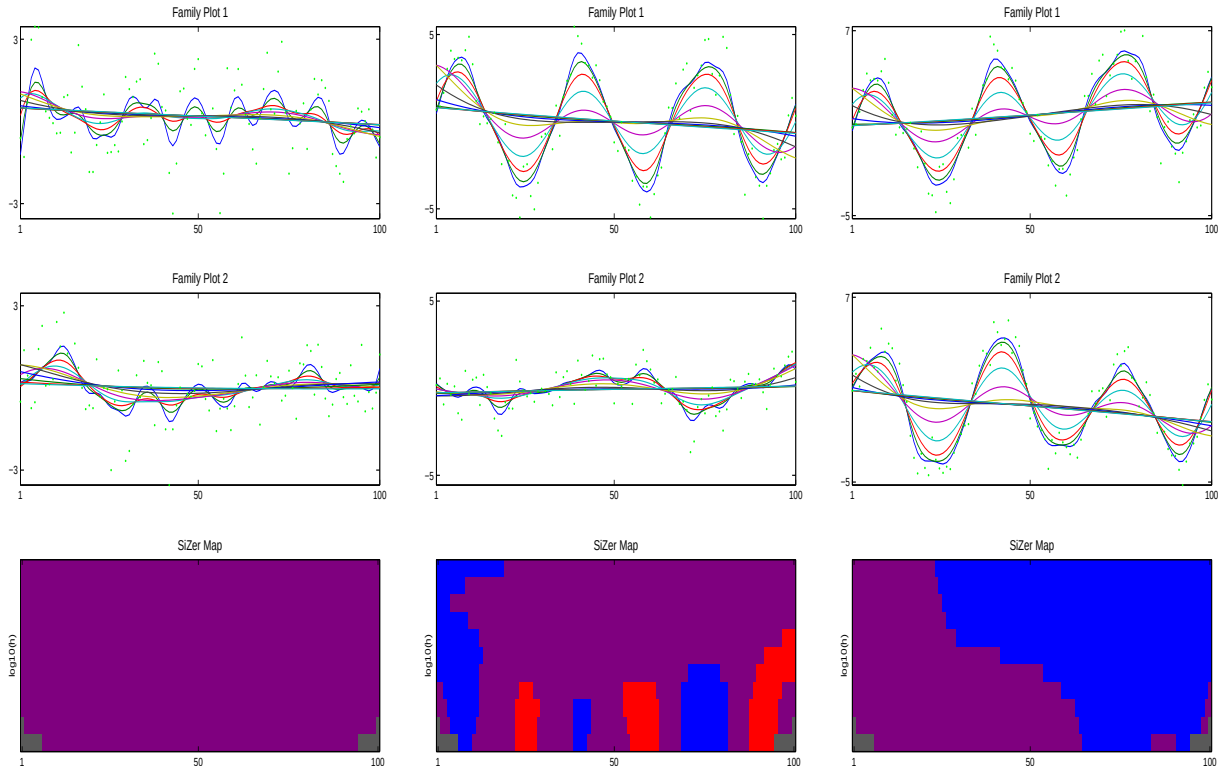
[24] Park, C., Hannig, J., and Kang, K.-H. (2009). Improved SiZer for Time Series. *Accepted, Statistica Sinica.*

[25] Park, C. and Kang, K. (2008). SiZer Analysis for the Comparison of Regression Curves. *Computational Statistics and Data Analysis*, **52**, 3954–3970.

[26] Park, C., Marron, J. S. and Rondonotti, V. (2004). Dependent SiZer: goodness of fit tests for time series models. *Journal of Applied Statistics*, **31**, 999–1017.

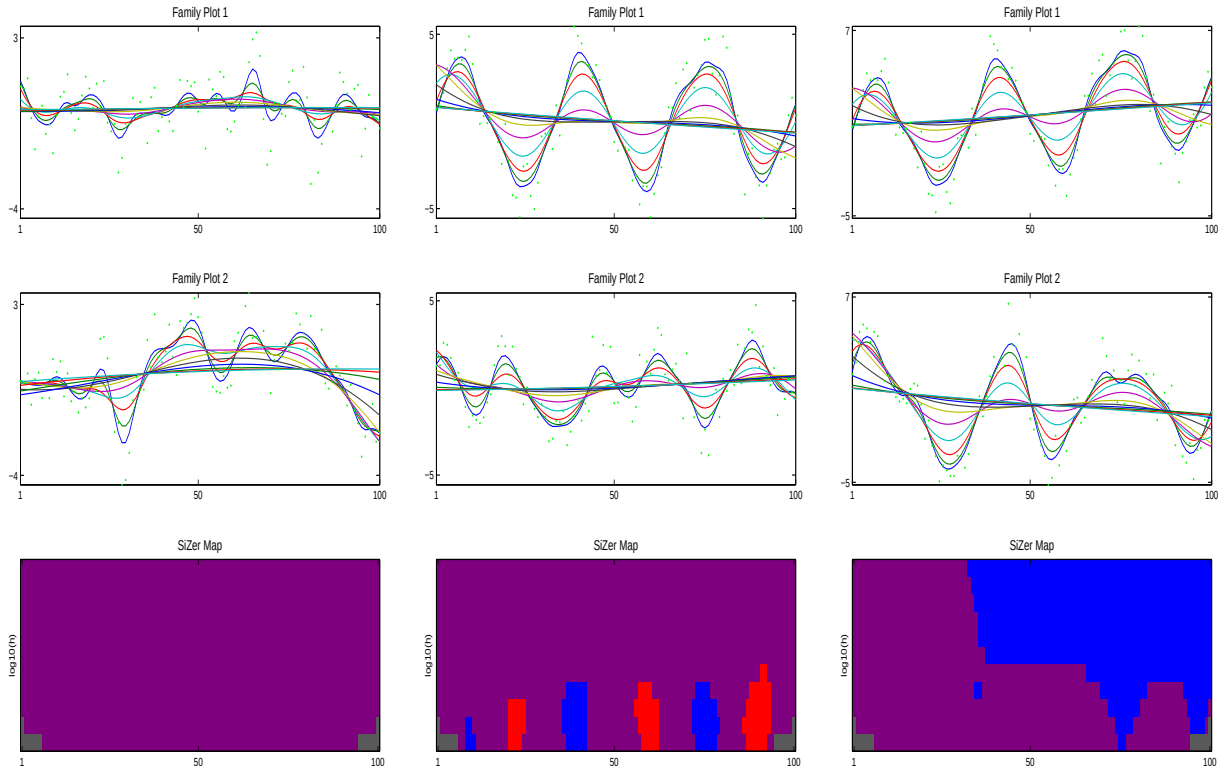
[27] Rondonotti, V., Marron, J. S. and Park, C. (2007). SiZer for time series: a new approach to the analysis of trends. *Electronic Journal of Statistics*, **1**, 268–289.

[28] Utev, S. A. (1990). The central limit theorem for $\varphi$-mixing arrays of random variables. *Theory of Probability and Its Applications*, **35**, 110–117.



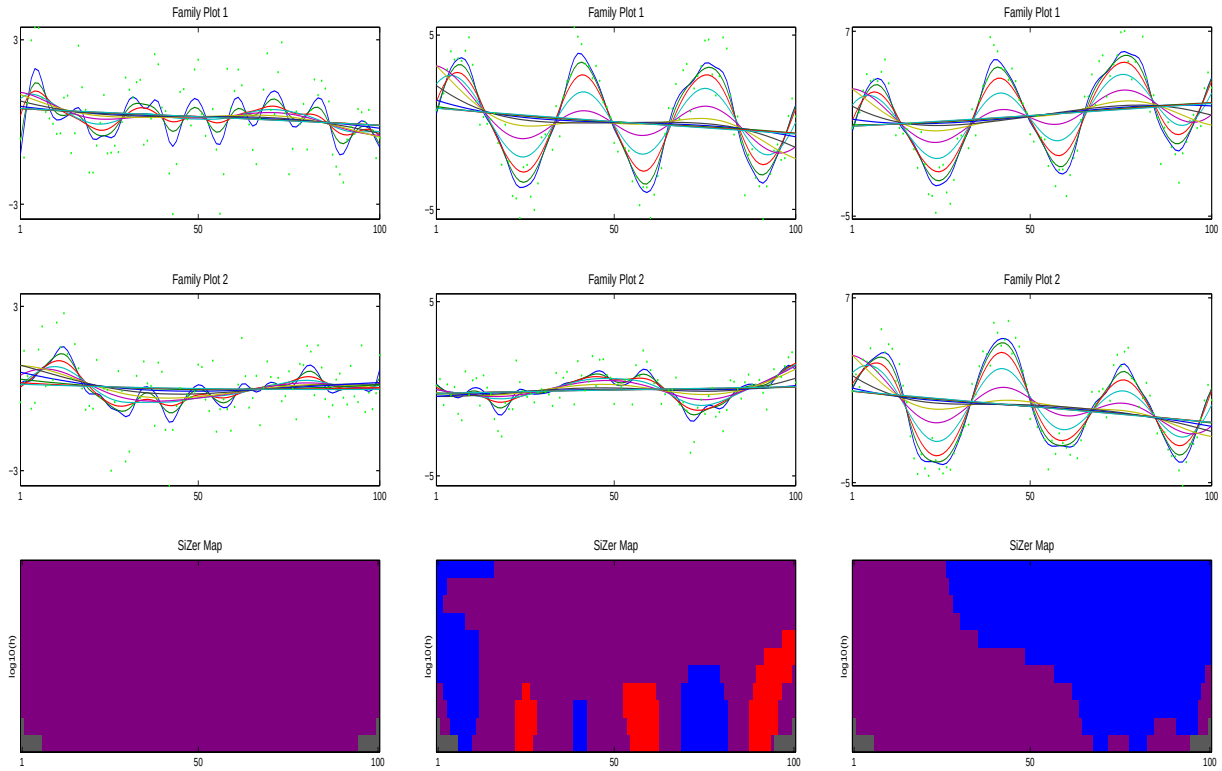

(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 1: Comparison of two time series with MA(1) and AR(1) errors. Autocovariance functions are given in advance.



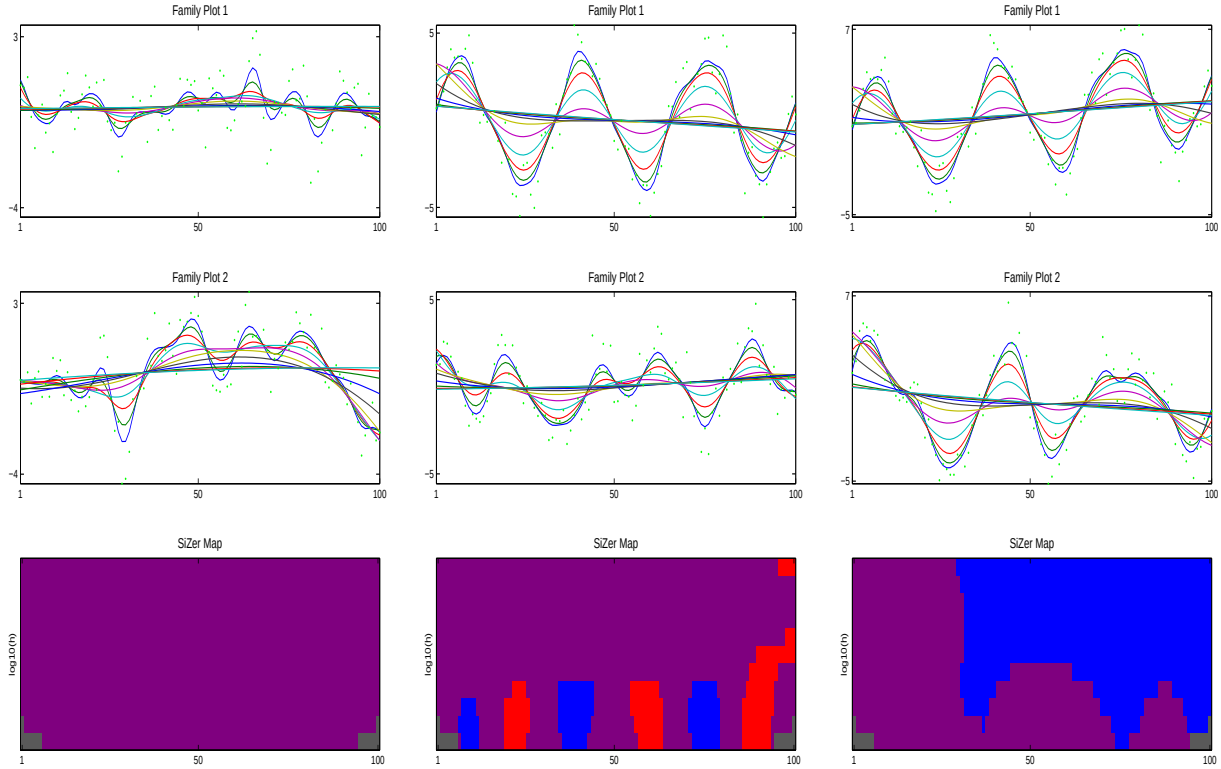

(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 2: Comparison of two time series with MA(1) and MA(5) errors. Autocovariance functions are given in advance.




(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3: Comparison of two time series with MA(1) and AR(1). Autocovariance functions are estimated from the time series.




(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 4: Comparison of two time series with MA(1) and MA(5). Autocovariance functions are estimated from the time series.

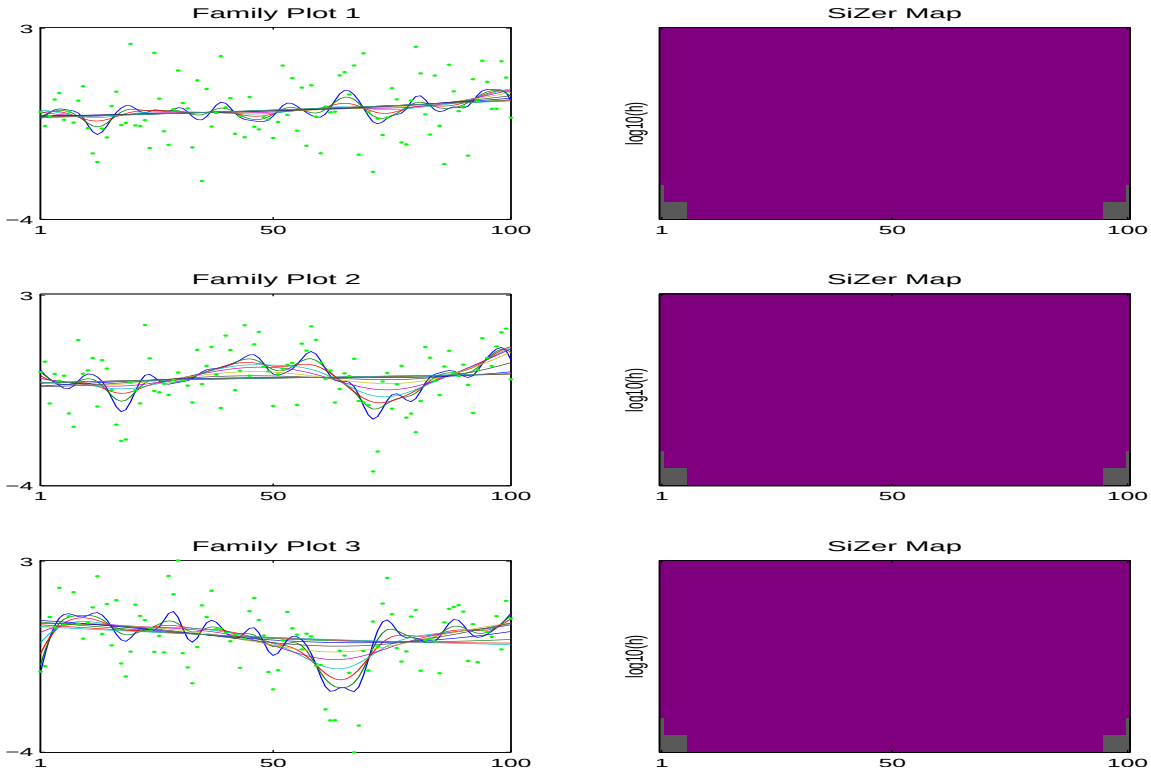



Figure 5: SiZer plots for comparing three time series with the same zero mean

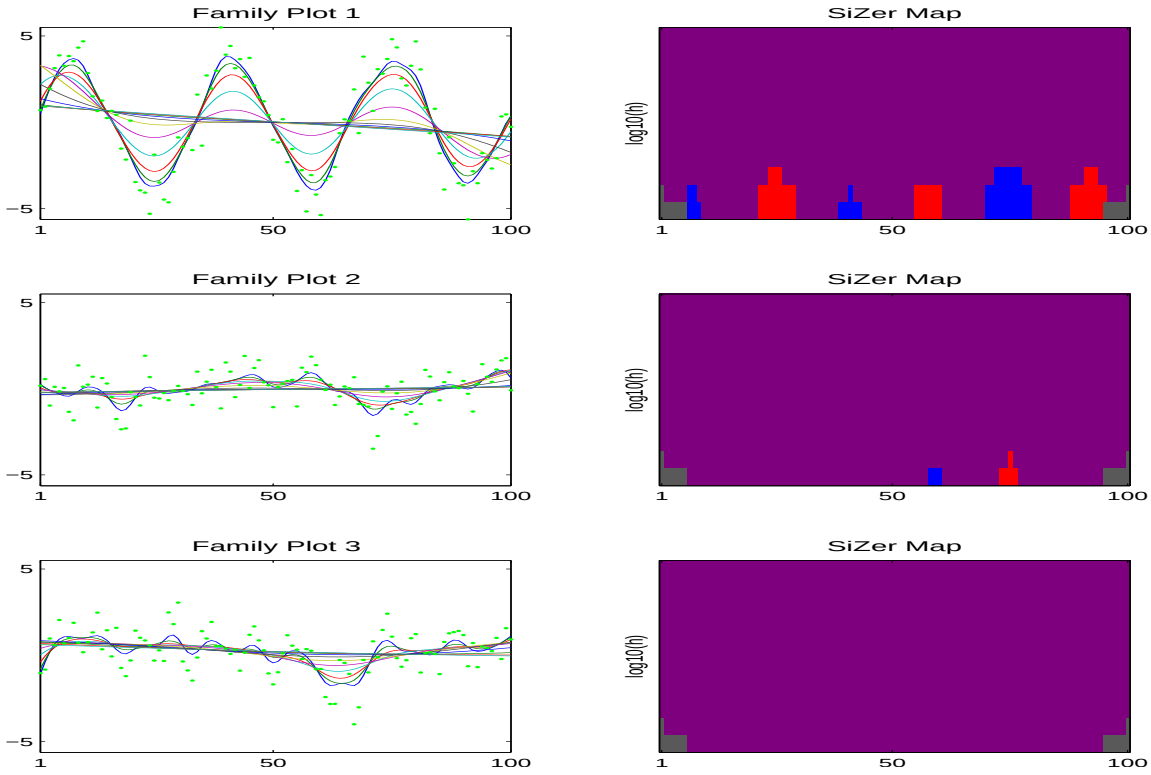
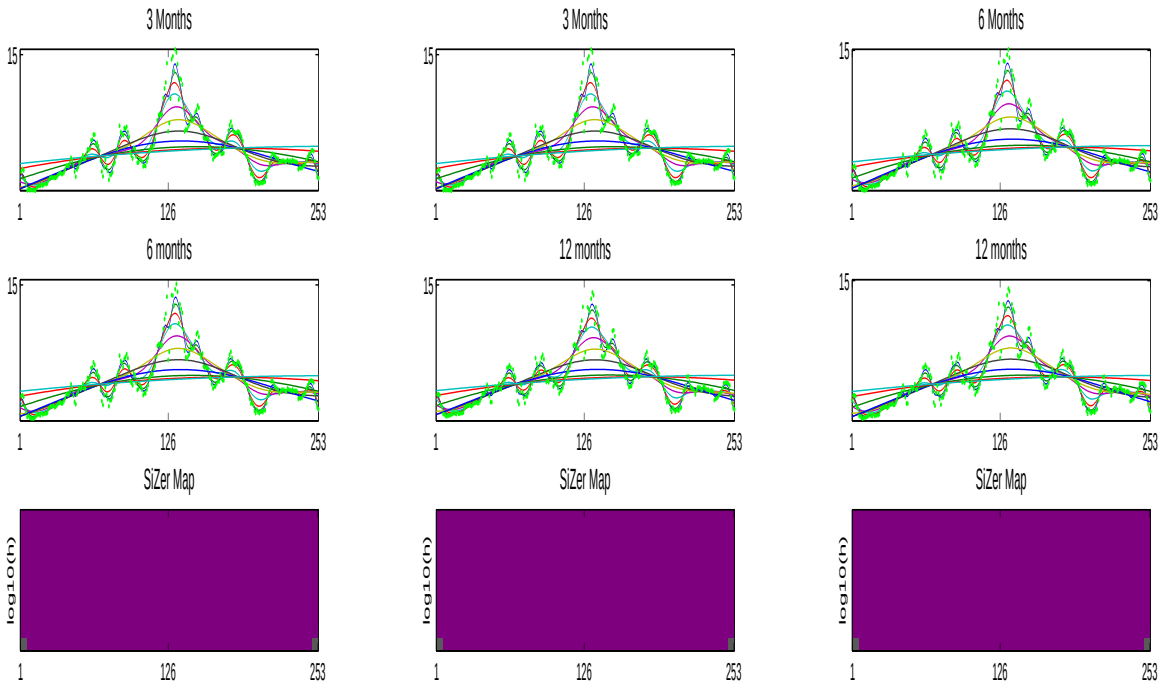
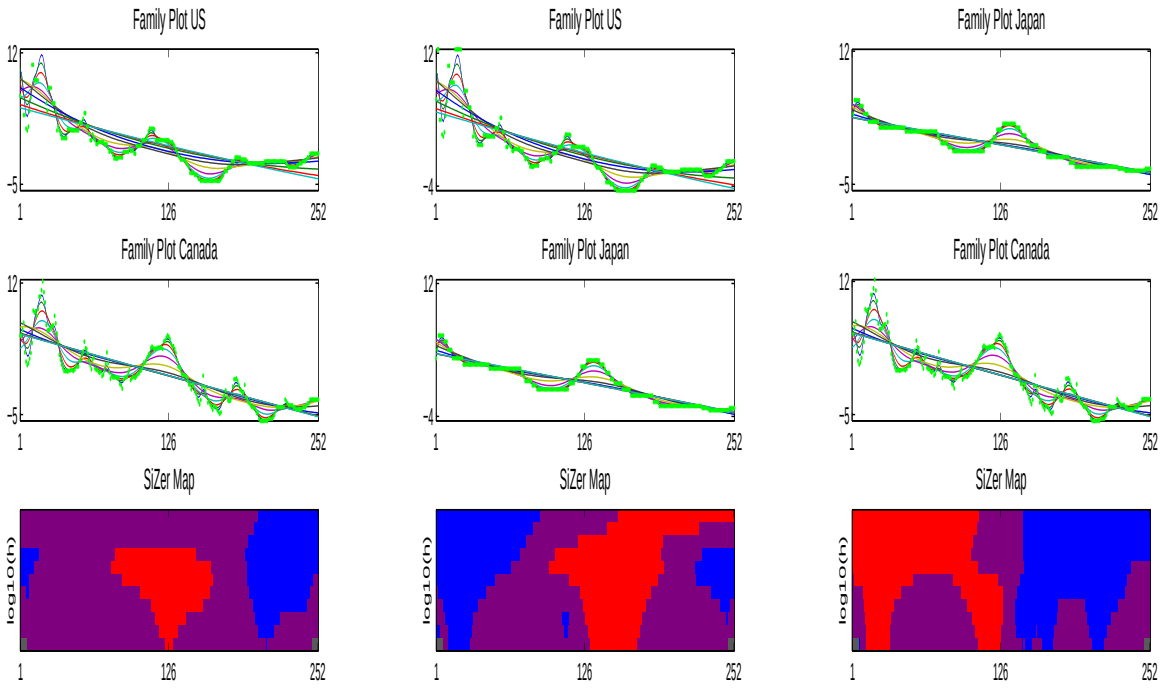



Figure 6: SiZer plots for comparing three time series with the different mean functions




(a) M3 vs. M6 (b) M3 vs. M12 (c) M6 vs. M12

Figure 7: Comparison of the yields of the 3-month, 6-month, and 12-month Treasury bills measured as the bi-monthly average from July 1959 to August 2001.




(a) US vs. Canada (b) US vs. Japan (c) Japan vs. Canada

Figure 8: Comparison of the trends for long term target interest rates for US, Canada, and Japan from January 1980 to December 2000.

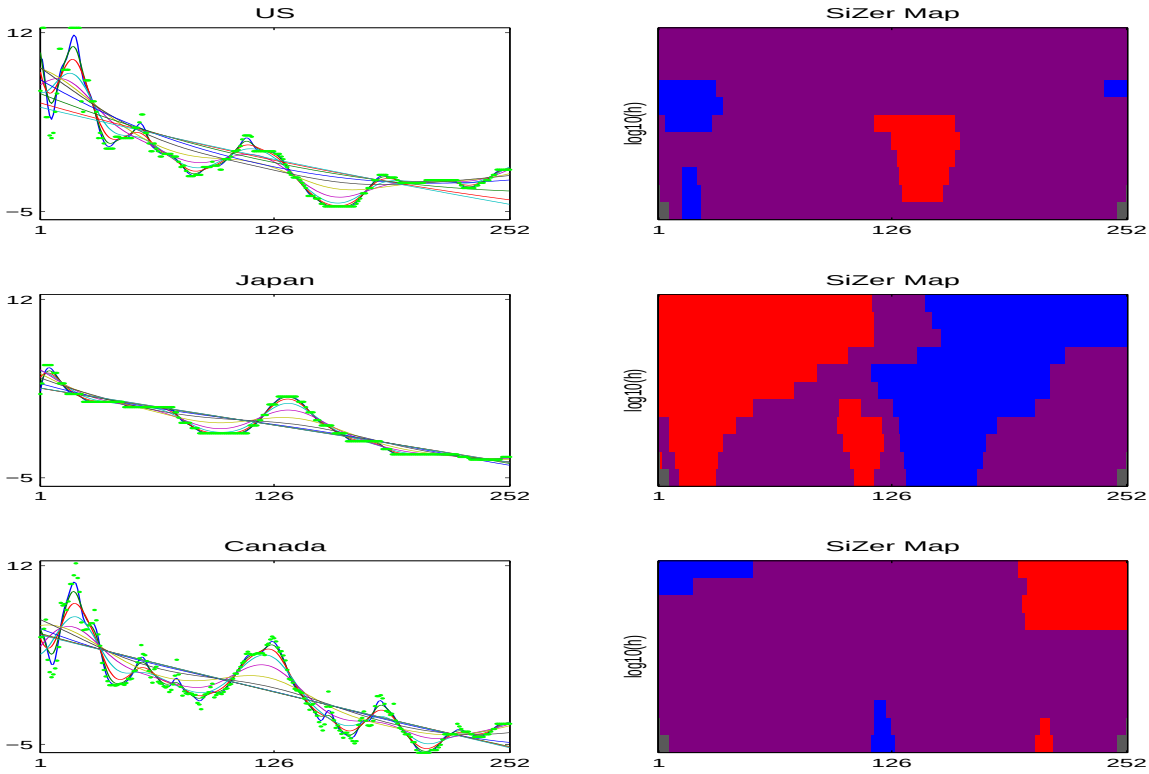



Figure 9: Multiple comparison of the trends for long term target interest rates for US, Canada, and Japan from January 1980 to December 2000.