Conditional Density Estimation and Density Forecast with Applications

by

Abdulaziz Dahir

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Probability and Statistics

Carleton University

Ottawa, Ontario

©2014

Abdulaziz Dahir
Abstract

In this thesis, several univariate and multivariate methods of building conditional density estimates and density forecasts are considered. For each method, two approaches (one-step and two-step) of finding the optimal bandwidth windows are used. To check which method is more accurate, numerous simulation studies were done on different univariate linear models such as $AR(1)$ and multivariate linear models such as $AR(2)$. Nonlinear models were also studied such as $AR(1)$ with ARCH errors. These ideas were later applied to real world data where the true densities are unknown and checked to see how good they forecast the next point. Stock market data was used where different methods of the multivariate one-step approach were applied. The stocks that were studied were the NYSE and TSX. The dependence of one stock market on the other was also studied and how much of a factor it plays in forecasting the next price. Weather data was also studied. The three weather variables studied were the daily averages of pressure, wind speed and temperature for the Ottawa region. Multiple forecast densities were built and compared to see which method has the highest accuracy. It was found that when dealing with non-linear data, the one-step method is more efficient in estimating the true conditional density. The two-step approach is better when trying to estimate a univariate conditional density. When trying to estimate multivariate conditional density the further one goes away from stationarity, the better that the 2-step approach is. Real world data indicates that multivariate methodology was better at predicting future events.
Acknowledgments

I would like to express the deepest appreciation to my supervisor Dr. Ould Haye, who has continuously helped me throughout my time as a graduate student. Without his help all of this would not have been possible so once again thank you Dr. Ould Haye. I would also like to thank all my committee members, Dr. Balan, Dr. Sinha and Dr. Villeneuve for taking the time to carefully read through my thesis and coming to my presentation. In addition, a thank you to Dr. Patterson for helping me obtain the weather data. Also, a thank you to Dr. Nash of University of Ottawa for passing on parts of his knowledge of the optimization function in R to me. A big thank you goes to my family who without their support along this stressful process of obtaining a Master’s degree would not have been possible. I can not forget to mention my office mates in Herzberg 3361 for their endless advice, aid and comfort throughout my time at the Carleton University campus. Finally, to those not mentioned here, just know that you are not forgotten and your help was really appreciated.
# Contents

1 Introduction ................................................. 4  
   1.1 Density Forecast ........................................ 4  

2 Univariate Case ............................................. 7  
   2.1 Joint Densities .......................................... 7  
   2.2 Conditional Densities ................................... 7  
   2.3 Conditional Density estimation ......................... 8  
   2.4 One Step Estimator ...................................... 9  
   2.5 Two Step Estimator ..................................... 10  
   2.6 Bandwidth Selection ................................... 12  
   2.7 Cross Validation ....................................... 12  
   2.8 Kullback-leibler Information Distance ................. 16  
   2.9 Nonlinear models ...................................... 17  
   2.10 Consistency of density estimates ..................... 19  
   2.11 Simulation ............................................ 26  

3 Multivariate Case ........................................... 38  
   3.1 One Step Estimator ..................................... 38  
   3.2 Two Step Estimator ..................................... 39  
   3.3 Bandwidth Selection ................................... 40
3.4 Multivariate Response ........................................... 41
3.5 Simulation ....................................................... 43

4 Multivariate Case With Real World Applications 51
4.1 Stock Market Data ............................................... 51
4.2 Weather Data .................................................... 57

5 Appendix .......................................................... 69
Chapter 1

Introduction

1.1 Density Forecast

A density forecast of a time series process at some future time is an estimate of the probability density function of the possible future values of that process. Due to this reason, density forecasts provide complete description of the uncertainty associated with a prediction, unlike point forecasts which contains no description of the associated uncertainty. Density forecasting is much closer to prediction intervals (P.I.) since P.I.s specify the probability that the actual outcome will fall within an interval. P.I.s are symmetric around the point forecast. Usually given as a band $\pm 1$ or $\pm 2$ standard errors around the point forecast and the associated statement of a probability of 68% or 95% with an assumption of a normal distribution. Occasionally Student’s distribution is used where the variance of the forecast error is estimated. At times there may be a departure from normality, still these calculations usually rest on a model with normally distributed innovations. Density forecasting can be seen as an alternative to point and interval forecasting approaches but little attention has been given to the evaluation of density forecasts. There are three key factors to explain the reason for this neglect. The first being that analytic construction of density forecasts has required restrictive, dubious assumptions, such as Gaussian innovations and no parameter estimation uncertainty. The use
of numerical and simulation techniques to build density forecasts has reduced the reliance on such assumptions. The second reason is that until recently there was little demand for density forecasts since point and interval forecasts were deemed efficient enough. Density forecast is mostly used in financial risk management. The final reason being the problem of density forecast evaluation appears difficult. Although there are methods developed for the evaluation of point and interval forecasts to the evaluation of density forecasts, but these methods lead to incomplete evaluation of density forecasts. In light of the increasing importance of density forecast as well as the lack of attention paid to it in the literature, which leads to the proposal of this thesis, which is to construct and evaluate density forecasts.

The first of the three factors above could be avoided by using nonparametric methods. In doing so there will be no need to check the usual assumptions of constant variance and the normality of errors. Also there is no need to check the distribution of the data or calculating the parameters of the distribution such as $\mu$ and $\sigma^2$ for the normal distribution. This is due to the fact that nonparametric methods do not rely on data belonging to any certain distribution thus there is no need to calculate the distribution parameters (see Brockwell and Davis (1991)). Conditional density estimation generalizes regression by modeling a full density $f(y \mid x)$ rather than only the expected value $E(y \mid x)$. This plays a major role in generating prediction intervals. This thesis investigates nonparametric kernel methods and their estimation of conditional density function. Conditional density is a ratio of the joint and marginal densities,

$$f(y \mid x) = \frac{f(x, y)}{f(x)}.$$  \(1.1\)

Thus the usual conditional density estimate is a ratio of joint density estimator and marginal density estimator

$$\hat{f}(y \mid x) = \frac{\hat{f}(x, y)}{\hat{f}(x)}. \quad (1.2)$$
The proposal of this thesis is to extend a two-stage estimator, developed by Hansen (2004) for one dimensional kernel conditional density estimation, to multivariate case and density forecast. The first stage is to estimate the conditional mean and to optimize its bandwidth to smooth the estimated conditional density. The second stage is to plug in the estimated conditional mean in estimating the conditional density. Since this method has two levels of smoothing it is expected that the curves of the two-stage estimate will be much smoother and reflect the properties of the true densities much better than the standard method of conditional density estimation technique.
Chapter 2

Univariate Case

2.1 Joint Densities

Let \((X_1, X_2)\) be a \(2 \times 1\) random vector with joint density \(f(x_1, x_2)\). This joint density satisfies

\[
P[a \leq X_1 \leq b, c \leq X_2 \leq d] = \int_a^b \int_c^d f(x_1, x_2) dx_1 dx_2. \tag{2.1}
\]

2.2 Conditional Densities

The conditional density of \(X_2\) given \(X_1\) can be calculated by the ratio of the joint density to the marginal density and is shown in the following formula

\[
f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f(x_1)} \tag{2.2}
\]

where the value of \(f(x_1)\) is fixed and greater than 0. The probability that \(X_2\) will fall between \(a\) and \(b\) given that \(X_1 = x_1\) is calculated by

\[
P[a \leq X_2 \leq b \mid X_1 = x_1] = \int_a^b f(x_2 \mid x_1) dx_2. \tag{2.3}
\]
2.3 Conditional Density estimation

There are two main methods to estimate densities which are

1. parametric

2. nonparametric

where the first approach assumes that the data is drawn from one of the known distributions. For example using the normal distribution the estimates of $\mu$ and $\sigma^2$ are calculated from the data and are substituted into the formula of the normal. This thesis examines nonparametric kernel methods for the estimation of the unknown density. It is assumed that the data comes from a distribution and has the unknown density $f$, only the data will be allowed to determine the estimate of the density. Unlike the case for parametric statistics where $f$ would be constrained to fall into a family of distributions, see Hansen (2004) and Silverman (1986).

The standard method of kernel density estimation is used. The kernel function is denoted by $K$ and satisfies the conditions

$$\int_{-\infty}^{\infty} K(x)dx = 1, \quad K(x) \geq 0 \quad \int_{-\infty}^{\infty} xK(x)dx = 0. \quad (2.4)$$

For the remainder of the thesis $K$ will be the standard normal density. Also a 2 step kernel density estimation method is used. The first step being to estimate the conditional mean and the second step being the estimation of the conditional density of the regression error. The 2 step estimator will require less smoothing, thus reducing the estimation variance, see Hansen (2004) and Silverman (1986).
2.4 One Step Estimator

Assume that one has a regression \( Y_i = m(X_i) + \epsilon_i \) where \( X_i \) is the explanatory variable, \( Y_i \) is the response variable and \( \epsilon_i \) is the regression error. The regression errors satisfy

\[
E(\epsilon_i) = 0 \quad \text{Var}(\epsilon_i) = \sigma^2 \quad \text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad \text{for} \quad i \neq j.
\]

The main objective is to estimate the conditional density of the random variable \( Y_i \) conditioned on \( X_i = x \), \( f(Y_i \mid X_i = x) \). The densities are estimated by the methods that are presented in Silverman (1986) which only deals with the marginal density. The ideas from Silverman (1986) are then extended to the conditional case as is stated in Hansen (2004). The standard method for kernel estimation of density of \( f(x) \) is

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) \tag{2.5}
\]

where \( K_h(u) = \frac{1}{h} K(u/h) \) and \( h = h(n) \) is the smoothing parameter which is also known as bandwidth, for simplicity it will be denoted only as \( h \). Based on a sample \((X,Y) = (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\) the standard kernel estimator of the joint density \( f(x,y) \) is

\[
\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^{n} K_{h_2}(x - X_i)K_{h_1}(y - Y_i). \tag{2.6}
\]

From (2.5) and (2.6) the standard kernel estimation of the conditional density \( f(y \mid x) \) is obtained from

\[
\hat{f}(y \mid x) = \frac{\hat{f}(x, y)}{\hat{f}(x)} \tag{2.7}
\]

\[
= \frac{\sum_{i=1}^{n} K_{h_2}(x - X_i)K_{h_1}(y - Y_i)}{\sum_{i=1}^{n} K_{h_2}(x - X_i)} \tag{2.8}
\]

(see Hansen (2004) and Silverman (1986)).
2.5 Two Step Estimator

For this method the conditional density is estimated in 2 steps as is stated in Hansen (2004). The conditional mean is calculated by the following formula

\[ m(x) = E(Y_i \mid X_i = x) \] (2.9)

for \( i = 1, 2, \ldots, n \) where \( Y_i = m(X_i) + \epsilon_i \) which is the same regression equation in the one step estimator method. The difference is estimating \( m(x) \) since \( m(x) \) is unknown, thus the first step is to estimate \( m(x) \) to achieve a final goal of attaining a conditional density of \( Y_i \) given \( X_i = x_i \). Nadayara (1964) and Watson (1964) proposed to estimate \( m \) as a locally weighted average using a kernel as a weighted function. The Nadayara-Watson estimator is

\[ \hat{m}(x) = \sum_{i=1}^{n} W_{b_0}(x, X_i) Y_i \] (2.10)

where the weight function \( W_{b_0}(x, y) = \frac{K_{b_0}(x-y)}{\sum_{i=1}^{n} K_{b_0}(x-y)} \) and \( b_0 \) is the bandwidth of the kernel function. The regression errors satisfy the same conditions in the one step estimator. The regression errors are estimated by

\[ \hat{\epsilon}_i = Y_i - \hat{m}(X_i). \] (2.11)

The Nadayara-Watson estimator (2.10) of \( m(x) \) is explained in terms of the estimated joint pdf \( \hat{f}(x, y) \) as follows

\[
\begin{align*}
m(x) &= E(Y_i \mid X_i = x) \\
&= \int_{-\infty}^{\infty} y f(y \mid x) dy \\
&= \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy}. \quad (2.12)
\end{align*}
\]
Dealing with the numerator first, substituting the joint density with its estimator leads to

\[
\int y \hat{f}(x, y) dy = \frac{1}{n} \int_{-\infty}^{\infty} y \left( \sum_{i=1}^{n} K_{h_2}(x - X_i) K_{h_1}(y - Y_i) \right) dy
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \left( K_{h_2}(x - X_i) \int_{-\infty}^{\infty} y K_{h_1}(y - Y_i) dy \right)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \left( K_{h_2}(x - X_i) \frac{1}{\sqrt{2\pi h_1}} \int_{-\infty}^{\infty} ye^{-\frac{(y-Y_i)^2}{2h_1^2}} dy \right)
\]

(2.13)

the last term of (2.13), \( \frac{1}{\sqrt{2\pi h_1}} \int_{-\infty}^{\infty} ye^{-\frac{(y-Y_i)^2}{2h_1^2}} \) is the expectation of the density \( N(Y_i, h_1^2) \) thus

\[
\frac{1}{\sqrt{2\pi h_1}} \int_{-\infty}^{\infty} ye^{-\frac{(y-Y_i)^2}{2h_1^2}} = Y_i
\]

(2.14)

This leads to numerator of the Nadaraya-Watson estimator of being \( \frac{1}{n} \sum_{i=1}^{n} K_{h_2}(x - X_i)Y_i \).

The denominator is much simpler where one only has to substitute (2.5) into the denominator of the Nadaraya-Watson estimator. Since \( h_2 \) is used to define the bandwidth for the one-step estimator of \( \hat{f}(y | x) \) we will convert it to \( b_0 \) which completes the explanation on how Nadaraya-Watson estimator is calculated, see Hansen (2004) and Silverman (1986).

Now moving onto the second step of the 2 step estimator method. Using the equation of (2.11) the term \( Y_i \) is replaced by \( \hat{e}_i + \hat{m}(X_i) \) this leads to the conditional density estimator of the two step method

\[
\hat{f}(y | x) = \frac{\sum_{i=1}^{n} K_{h_2}(x - X_i) K_{h_1}(y - \hat{m}(x) - \hat{e}_i)}{\sum_{i=1}^{n} K_{h_2}(x - X_i)}.
\]

(2.15)

2.6 Bandwidth Selection

Many measures have been used to study the distance of the density estimator \( \hat{f}(y|x) \) from the true density \( f(y|x) \). One of the most commonly used measure is the mean square error (MSE) which is defined by

\[
MSE_x(\hat{f}(y | x)) = E(\hat{f}(y | x) - f(y | x))^2. \tag{2.16}
\]

Another method is the integrated square error (ISE) where it was defined in Silverman (1986) for the marginal density case and was extended to the conditional density case by Hansen (2004), defined as

\[
I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\hat{f}(y | x) - f(y | x))^2 f(x)dydx
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(y | x)^2 f(x)dydx - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(y | x)f(y | x)f(x)dydx
\]

\[
+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y | x)^2 f(x)dydx
\]

\[
= I_1 - 2I_2 + I_3. \tag{2.17}
\]

Since the bias of the estimate is only dependent on the bandwidth, the term \( I_3 \) does not depend on the bandwidth thus, \( I_3 \) term could be ignored. The concept of ISE is to pick values for the bandwidth parameters that would result in the estimate of \( I \) to be as close to 0 as possible.

2.7 Cross Validation

In the real world the ISE method is regarded as useless since the true densities \( f(x) \) and \( f(y|x) \), are never known and if they are known then estimating them is pointless. Since \( I_1 \) and \( I_2 \) both contain true densities in them then \( I_1 \) and \( I_2 \) both need to be estimated. The so-called leave-one-out-cross-validation methodology is applied to estimate \( I_1 \) and \( I_2 \). Let \( \tilde{f}_{-i}(y | x) \)
denote the one step estimator conditional density with observation \(i\) omitted i.e.,

\[
\tilde{f}_{-i}(Y_i \mid X_i) = \frac{\sum_{i \neq j} K_{h_2}(X_i - X_j)K_{h_1}(Y_i - Y_j)}{\sum_{i \neq j} K_{h_2}(X_i - X_j)}.
\] (2.18)

The cross validation function is defined by Hansen (2004) as

\[
\tilde{I} = \tilde{I}_1 - 2\tilde{I}_2.
\] (2.19)

The basic principle of leave one out cross validation is to construct an estimate of the \(I\) from the data and then minimize this estimate over the bandwidth parameters to give an optimal value of these parameters. The estimate of the leave one out conditional density takes into account all the data points except \(X_i\). The estimate of \(I_1\) for the one step procedure is

\[
\tilde{I}_1 = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \tilde{f}_{-i}(y \mid X_i)^2 dy
\] (2.20)

Expanding (2.20) leads to

\[
\tilde{I}_1 = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left( \frac{\sum_{j \neq i} K_{h_2}(X_i - X_j)K_{h_1}(y - Y_j)}{\sum_{j \neq i} K_{h_2}(X_i - X_j)} \right)^2 dy
\] (2.21)

\[
= (1/n) \sum_{i=1}^{n} \sum_{j \neq i} K_{h_2}(X_i - X_j) \sum_{k \neq i} K_{h_2}(X_i - X_k) \left( \frac{\sum_{j \neq i} K_{h_2}(X_i - X_j)}{(\sum_{j \neq i} K_{h_2}(X_i - X_j))^2} \right) \int_{-\infty}^{\infty} K_{h_1}(y - Y_j)K_{h_1}(y - Y_k)dy.
\]
Now dealing with the integrand portion of (2.21)

\[
\int_{-\infty}^{\infty} K_{h_1}(y - Y_j) K_{h_1}(y - Y_k) dy = \frac{1}{2\pi h_1^2} \int_{-\infty}^{\infty} e^{-\frac{(y-y_j)^2 + (y-y_k)^2}{2h_1^2}} dy
\] 

(2.22)

\[
= \frac{1}{2\pi h_1^2} \int_{-\infty}^{\infty} e^{-\frac{(y^2-2yY_j-Y_j^2+Y_k^2+Y_k^2)}{2h_1^2}} dy
\]

\[
= \frac{1}{2\pi h_1^2} e^{-\frac{(y^2-Y_j^2)}{2h_1^2}} \int_{-\infty}^{\infty} e^{-\frac{(y-y_j+Y_j)}{h_1^2}} dy
\]

\[
= \frac{1}{2\pi h_1^2} e^{-\frac{(y^2+Y_j^2)}{2h_1^2}} \int_{-\infty}^{\infty} e^{-\frac{(y-2y(Y_j+Y_k))}{h_1^2}} dy
\]

\[
= \frac{1}{\sqrt{2\pi} \sqrt{2h_1}} e^{-\frac{(Y_k-Y_j)^2}{4h_1^2}} \frac{1}{\sqrt{2\pi} h_1^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{(y-2y(Y_j+Y_k))}{h_1^2}} dy
\]

the last line of (2.22) the second part of that line is the density of \( N(\frac{Y_j+Y_k}{2}, \frac{h_1^2}{2}) \). We are left with

\[
\frac{1}{\sqrt{2\pi} \sqrt{2h_1}} \frac{1}{\sqrt{2\pi} h_1^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{(y-2y(Y_j+Y_k))}{h_1^2}} dy = \frac{1}{\sqrt{4\pi h_1}} e^{-\frac{(Y_k-Y_j)^2}{4h_1^2}}
\]

\[
= K_{\sqrt{2h_1}}(Y_k - Y_j).
\] 

(2.23)

Finally the estimate of \( I_1 \) is

\[
\tilde{I}_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} K_{h_2}(X_i - X_j) \sum_{k \neq i} K_{h_2}(X_i - X_k) K_{\sqrt{2h_1}}(Y_k - Y_j)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} K_{h_2}(X_i - X_j) \sum_{k \neq i} K_{h_2}(X_i - X_k) K_{\sqrt{2h_1}}(Y_k - Y_j)
\]

(2.24)

The estimate for \( I_2 \) is much more straightforward since its just the estimate of the condi-
tional density of \( f(Y_i \mid X_i) \) with the value of \( X_i \) being left out

\[
\tilde{I}_2 = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_{-i}(Y_i \mid X_i)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} K_{b_2}(X_i - X_j) \frac{K_{b_1}(Y_i - Y_j)}{\sum_{j \neq i} K_{b_2}(X_i - X_j)}
\] (2.25)

For the 2 step estimator, the optimal bandwidths are obtained via 2 steps, see Hansen (2004). The bandwidth of the Nadaraya-Watson estimator is found from

\[
\hat{m}_{-i}(X_i) = \sum_{j \neq i} W_{b_0}(X_i, X_j) Y_j
\] (2.26)

where \( W_{b_0}(X_i, X_j) = \frac{K_{b_0}(X_i - X_j)}{\sum_{i \neq j} K_{b_0}(X_i - X_j)} \). All the values are used again except for \( X_i \). The value of the function of \( \hat{m}(X_i) \) is then used in estimating the residuals (2.11). The least square error of the residuals is calculated

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{e}_i)^2.
\] (2.27)

The MSE function is then optimized to get the value of the bandwidth \( b_0 \) to give the smallest value of the MSE.

Letting \( \hat{f}_{-i}(y \mid x) \) be the cross validation estimate of the conditional density using the 2 step method and leaving the observation \( X_i \) out of the calculations. The 2 step cross validation function is defined by Hansen (2004) as \( \hat{I} = \hat{I}_1 - 2\hat{I}_2 \) where

\[
\hat{I}_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} K_{b_2}(X_i - X_j) \frac{K_{b_2}(X_i - X_k)K_{b_1}(y - \hat{m}(X_j) - \hat{e}_i)}{\left(\sum_{j \neq i} K_{b_2}(X_i - X_j)\right)^2}.
\] (2.28)

The estimate of \( I_2 \) using the 2 step method is

\[
\hat{I}_2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} K_{b_2}(X_i - X_j) \frac{K_{b_1}(Y_i - \hat{m}(X_j) - \hat{e}_i)}{\sum_{j \neq i} K_{b_2}(X_i - X_j)}.
\] (2.29)
It has been proven by Stone (see Stone (1984)) for a very large sample that the cross validation function $\hat{I}$ and $\tilde{I}$ are consistent estimates of the oracle function $I$. By oracle what is meant is that the true density is known. We have $\frac{\hat{I}}{I} \to 1$ as $n \to \infty$ in probability (see Stone (1984)). Therefore asymptotically, the leave one out cross validation method achieves the best possible choice for the bandwidth variables. Simply put mathematically the bandwidth parameters found using the cross validation method are as optimal as those of oracle, therefore there is no need to know the actual true density since the leave one out method estimates the true density very accurately.

2.8 Kullback-leibler Information Distance

The Kullback-leibler information distance between the conditional density $f(y \mid x)$ and its estimator $\hat{f}$ is defined by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(y \mid x)}{\hat{f}(y \mid x)}\right) dy dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log(f(y \mid x)) dy dx - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log(\hat{f}(y \mid x)) dy dx.$$

Since the first part of (2.30) does not depend on the bandwidth parameter it can be ignored. Since the only term left is $-\int_{-\infty}^{\infty} f(x) \log(\hat{f}(y \mid x))$ thus minimizing it yields a density that is close to the true one. The two-step procedure follows the same principles. Kullback-leibler information is also known as Kullback-Leibler divergence, information divergence, information gain and relative entropy (see Silverman (1986)). For the simulations this methodology is efficient since the distance between the true and estimated densities could be measured to see how close they are. Here again since the marginal and conditional densities are unknown we will have to use a cross validation method to obtain optimal bandwidth from this log likelihood.
method.

\[
L = \frac{1}{n} \sum_{i=1}^{n} \log(\tilde{f}_{-i}(Y_i | X_i) \tilde{f}_{-i}(X_i)) = \frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{\sum_{j \neq i} K_h_1(Y_i - Y_j) K_h_2(||X_i - X_j||)}{\sum_{j \neq i} K_h_2(||X_i - X_j||)}\right) + \log\left(\frac{1}{n - 1} \sum_{j \neq i} K_h_2(||X_i - X_j||)\right) = \frac{1}{n} \sum_{i=1}^{n} \log\left(\sum_{j \neq i} K_h_1(Y_i - Y_j) K_h_2(||X_i - X_j||)\right) - \log(n - 1). \tag{2.31}
\]

Since the log likelihood function is double nested with summation unlike the ISE cross validation (CV) function which was triply nested under summation thus making log likelihood function much quicker to compute and has computation time of \(O(n^2)\) (Holmes et al (2007) and Silverman (1986)) compared to \(O(n^3)\) of the ISE CV method.

### 2.9 Nonlinear models

Recent problems in the financial world have motivated the study of volatility of a time series. A setback of the linear stationary time series models is their failure to account for change in volatility. An example of volatility in a time series is a strong negative score followed by a period or score of a high value. Although ARMA models assume constant variance, models such as the autoregressive conditionally heteroscedastic (ARCH) were developed to model the changes in volatility. In an ARCH model it is not the variance itself that changes but the conditional variance in a specific way depending on the data. Looking at an AR(1) model \(x_t = \phi x_{t-1} + \epsilon_t\) where \(\epsilon_t\) are i.i.d. with mean 0 and constant variance. If one wants to predict \(x_t\) then since it is an AR(1) model then it is only dependent on the one previous score. Due to this, the best predictor of \(x_t\) is the conditional mean, \(E(x_t \mid x_{t-1})\). The success of linear models such as AR(1) model for forecasting purposes is due to the fact that the conditional mean is dependent on the most recent data point and evolve with time. The conditional variance of an i.i.d. white noise \(\epsilon_t\) is \(\text{var}(\epsilon_t \mid \epsilon_{t-1}) = \text{var}(\epsilon_t) = \sigma^2\) which is constant since it does not depend
on $t$. In the ARCH model the conditional variance is dependent on time. The conditional variance is what quantifies our uncertainty about the future observation. For example if $x_t$ is the value of a stock at time $t$ then we define its return $y_t$ as

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}}. \quad (2.32)$$

Often financial time series exhibit return that do not have constant conditional variance. Highly volatile periods tend to be clustered together. Such behavior can be modeled with the simplest ARCH model which is ARCH(1) where the return is modeled as

$$y_t = \sigma_t \epsilon_t \quad (2.33)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 \quad (2.34)$$

where $\epsilon_t$ is standard Gaussian white noise $\epsilon_t \sim N(0, 1)$ also $\alpha_1 > 0$. The conditional distribution of the return given one time previous is Gaussian

$$y_t \mid y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2). \quad (2.35)$$

From here one can see that $Var(y_t \mid y_{t-1}) = \alpha_0 + \alpha_1 y_{t-1}^2$ which is not constant over time. Also it is possible to write the ARCH(1) model as a non-Gaussian AR(1) where the returns $y_t$ are squared, $y_t^2$. Now (2.33) becomes

$$y_t^2 = \sigma_t^2 \epsilon_t^2 \quad (2.36)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 \quad (2.37)$$

and adding the two equations above we get

$$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \nu_t \quad (2.38)$$
where \( v_t = \sigma^2_t (\epsilon_t^2 - 1) \) (see for example Shumway (2011)).

**Remark** When it comes to point forecast or prediction intervals there is no actual difference between modeling with independent errors and ARCH(1) errors. The difference is important only when it comes to conditional prediction intervals. In this case, modeling with ARCH(1) errors allows for more accurate intervals (see Churvich’s notes).

### 2.10 Consistency of density estimates

In nonparametric density estimations the estimated density are created from a dataset with \( n \) observations. It was previously stated that the bandwidth parameter \( h \) is somewhat explicitly dependent on the sample size \( n \). Limiting results are then obtained on the behavior of the conditional density estimate as \( n \to \infty \). To show the explicit dependence of the bandwidth parameter on \( n \) for this section it will be written as \( h_n \). Consistency of the estimate \( f \) at a point \( y \) given \( x \) was studied and under assumptions on the kernel \( K \) where that \( K \) is a bounded Borel function satisfying

\[
\int_{-\infty}^{\infty} K(t) dt = 1 \tag{2.39}
\]

also

\[
| tK(t) | \to 0 \quad \text{as} \quad | t | \to \infty \tag{2.40}
\]

and

\[
h_n \to 0, \quad nh_n^2 \to \infty. \tag{2.41}
\]

as shown in Parzen (1979) and Silverman (1986).
Since \( f(y \mid x) = \frac{f(x,y)}{f(x)} \) dealing with estimated conditional density being a consistent estimator is dealing with the numerator and denominator separately. For the numerator, the joint density case it was proven in Rosenblatt (1969) for the i.i.d. case but this result will be extended to the dependent case. The kernel joint density can be written as

\[
\tilde{f}_n(x, y) = \frac{1}{n b_n^2} \sum_{i=1}^{n} h\left(\frac{x-X_i}{b_n}, \frac{y-Y_i}{b_n}\right).
\] (2.42)

An estimate is consistent if it is asymptotically unbiased and its variance goes to 0. First dealing with the bias problem where bias is \( E(\tilde{f}_n(x, y) - f(x, y) \) where

\[
E(\tilde{f}_n(x, y)) = E\left(\frac{1}{n b_n^2} \sum_{i=1}^{n} h\left(\frac{x-X_i}{b_n}, \frac{y-Y_i}{b_n}\right)\right)
= \frac{1}{b_n^2} E(h(\frac{x-X_i}{b_n}, \frac{y-Y_i}{b_n}))
= \frac{1}{b_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\frac{x-u}{b_n}, \frac{y-v}{b_n}) f(u, v) du dv.
\] (2.43)

The next step being a change of variables where letting \( \alpha = \frac{x-u}{b_n} \) and \( \beta = \frac{y-v}{b_n} \) which leads to \( u = x-b\alpha, \quad du = -bd\alpha, v = y-b\beta \) and \( dv = -bd\beta \). Substituting these new variables back into (2.42) leads to

\[
\frac{1}{b_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) f(x-b_n\alpha, y-b_n\beta) b_n^2 d\alpha d\beta
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) f(x-b_n\alpha, y-b_n\beta) d\alpha d\beta
\] (2.44)

using Taylor’s expansion to the first order on the function \( f(x-b_n\alpha, y-b_n\beta) \)

\[
f(x-b_n\alpha, y-b_n\beta) \approx f(x, y) + b_n\alpha f_x(x, y) + b_n\beta f_y(x, y) + o(b_n).
\] (2.45)
Substituting (2.45) back into (2.44)

\[
f(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) d\alpha d\beta + b_n f_x(x, y) \int_{-\infty}^{\infty} \alpha d\alpha \int_{-\infty}^{\infty} h(\alpha, \beta) d\alpha d\beta
\]

\[
+ b_n f_y(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \beta h(\alpha, \beta) d\alpha d\beta + o(b_n) = f(x, y) + o(b_n)
\]

(2.46)

where \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) d\alpha d\beta = 1 \) since it is a density and \( \int_{-\infty}^{\infty} \alpha h(\alpha, \beta) d\alpha = \int_{-\infty}^{\infty} \beta h(\alpha, \beta) d\beta = 0 \) as is shown in Rosenblatt (1969). This all leads to

\[
Bias = f(x, y) + o(b_n) - f(x, y)
\]

\[
= o(b_n)
\]

(2.47)

which proves that the numerator of the conditional density estimate is an asymptotically unbiased estimator of the joint density.

Now dealing with the variance of the joint density a quick example of the approach taken is

\[
V \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} V \left( \sum_{i=1}^{n} X_i \right)
\]

\[
= \frac{1}{n} V(X_1) + \frac{1}{n^2} \sum_{i \neq j} \sum Cov(X_i, X_j).
\]

(2.48)

The previous equations is derived from the fact that \( V(X_i) = V(X_1) \). Now going back to \( V(\hat{f}_n(x, y)) \)

\[
V(\hat{f}_n(x, y)) = V \left( \frac{1}{nb_n^2} \sum_{i=1}^{n} h \left( \frac{x - X_i}{b_n}, \frac{y - Y_i}{b_n} \right) \right)
\]

\[
= \frac{1}{nb_n^4} V(h \left( \frac{x - X_i}{b_n}, y - Y_i b_n \right))
\]

\[
+ \frac{2}{n^2 b_n^4} \sum_{j=2}^{n} \sum_{i=1}^{j-1} Cov \left( h \left( \frac{x - X_i}{b_n}, \frac{y - Y_i}{b_n} \right), h \left( \frac{x - X_j}{b_n}, \frac{y - Y_j}{b_n} \right) \right)
\]

(2.49)
where the first term of the last line of (2.49) is proven to be 0 in Rosenblatt (1969). Dealing only with covariance term now, let \( Z_i = h(\frac{x-X_i}{b_n}, \frac{y-Y_i}{b_n}) \) and \( Z_j = h(\frac{x-X_j}{b_n}, \frac{y-Y_j}{b_n}) \). Since \( Z_i \) is stationary thus \( \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_1, Z_{j-i+1}) = \gamma(j - i) \) letting \( j - i = t \) thus \( j = t + i \) leads to \( 2 \leq t + i \leq n \) resulting to \( 2 - t \leq i \leq n - t \). The covariance term is

\[
\frac{2}{n^2 b_n^4} \sum_{t=1}^{n-1} \sum_{i=1}^{n-t} \gamma(t) = \frac{2}{n^2 b_n^4} \sum_{t=1}^{n} (n-t)\gamma(t) = \frac{2}{nb_n^4} \sum_{t=1}^{n} \frac{n-t}{n} \gamma(t). \quad (2.50)
\]

Since \( \frac{n-t}{n} \leq 1 \) all that is left to focus on is \( \gamma(t) \)

\[
\gamma(t) = \text{Cov}(Z_1, Z_{1+t}) = E(Z_1Z_{1+t}) - E(Z_1)E(Z_{1+t}) = E(Z_1Z_{1+t}) - (E(Z_1))^2 \quad (2.51)
\]

the last line comes about the fact again \( Z_1 \) is stationary thus \( E(Z_1) = E(Z_{1+t}) \). Dealing with the first term of \( \gamma(t) \)

\[
E(Z_1Z_{1+t}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\frac{x-u}{b_n}, \frac{y-v}{b_n})h(\frac{x-u'}{b_n}, \frac{y-v'}{b_n}) f(x_1, y_1, x_{1+t}, y_{1+t}) (u, v, u', v') dudvda'db'd
\]

\[
= b_n^4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta)h(\alpha', \beta') f(x_1, y_1, x_{1+t}, y_{1+t}) (x - b_n\alpha, y - b_n\beta, x - b_n\alpha', y - b_n\beta') d\alpha d\beta a'db'd \quad (2.52)
\]

again using the first order Taylor’s expansion eventually leads to \( f(z_1, z_{1+t})(x, y, x, y) + o(b_n) \).

The second term of the cross term \( (E(Z_1))^2 = (f_{Z_1}(x, y))^2 \) which finally leads to

\[
\frac{2}{n} \sum_{t=1}^{n} f_{z_1, z_{1+t}}(x, y, x, y) + o(b_n) - (f_{Z_1}(x, y))^2 \quad (2.53)
\]

as \( t \to \infty \) we get \( f_{z_1, z_{1+t}} \approx f_{z_1} f_{z_{1+t}} = (f_{Z_1})^2 \) this property is shown in Giraitis and Sur-
gailis (2003) finally getting the result $\gamma(t) = o(b_n)$ where $b_n \to 0$ which means $V(\tilde{f}_n(x,y))$ is asymptotically 0.

The denominator of the conditional density is just the marginal density. The estimated marginal density is

$$\tilde{f}_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} h \left( \frac{x - X_i}{b_n} \right). \quad (2.54)$$

Its variance is estimated as

$$Var(\tilde{f}_n(x)) = V \left( \frac{1}{nb_n} \sum_{i=1}^{n} h \left( \frac{x - X_i}{b_n} \right) \right)$$

$$= \frac{1}{nb_n^2} V \left( h \left( \frac{x - X_i}{b_n} \right) \right) + \frac{2}{n^2b_n^2} \sum_{j=2}^{n} \sum_{i=1}^{j-1} Cov \left( h \left( \frac{x - X_i}{b_n} \right), h \left( \frac{x - X_j}{b_n} \right) \right) \quad (2.55)$$

where the first term on the last line of (2.55) goes to 0 as was shown in the joint density case.

Letting $z_i = h \left( \frac{x - X_i}{b_n} \right), z_j = h \left( \frac{x - X_j}{b_n} \right)$ where $z_i$ is stationary, leads to

$$\frac{2}{n^2b_n^2} \sum_{j=2}^{n} \sum_{i=1}^{j-1} Cov(z_i, z_j) = \frac{2}{n^2b_n^2} \sum_{j=2}^{n} \sum_{i=1}^{j-1} Cov(z_1, z_{j-i+1}). \quad (2.56)$$

Letting $t = j - i$ results in

$$\frac{2}{n^2b_n^2} \sum_{t=1}^{n} \sum_{i=1}^{t-1} \gamma(t) = \frac{2}{n^2b_n^2} \sum_{t=1}^{n} \left( \frac{n - t}{n} \right) \gamma(t) \quad (2.57)$$

since $\left( \frac{n-t}{n} \right) \leq 1$ therefore it is considered negligible and can be ignored.

$$\gamma(t) = Cov(z_1, z_{t+1})$$

$$= E(z_1 z_{t+1}) - E(z_1)E(z_{t+1})$$

$$= E(z_1 z_{t+1}) - (E(z_1))^2 \quad (2.58)$$
$$E(z_1 z_{t+1}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h\left(\frac{x-u}{b_n}\right) h\left(\frac{x-u'}{b_n}\right) f_{X_1, X_{1+t}}(u, u') dudu'$$

$$= b_n^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha) h(\alpha') f(x - b_n \alpha, x - b_n \alpha') d\alpha d\alpha'$$

(2.59)

A variable change has occurred where $\alpha = \left(\frac{x-u}{b_n}\right)$ and $\alpha' = \left(\frac{x-u'}{b_n}\right)$. After performing a first order Taylor series expansion on $f(x - b_n \alpha, x - b_n \alpha')$ results in

$$f(x - b_n \alpha, x - b_n \alpha') = f(x, x) + b_n \alpha f(x, x) + b_n \alpha' f(x, x) + o(b_n)$$

(2.60)

thus,

$$E(z_1 z_{t+1}) = b_n^2 f(x, x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha) h(\alpha')$$

$$+ b_n^2 f(x, x) b_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha h(\alpha) h(\alpha') d\alpha d\alpha'$$

$$+ b_n^2 f(x, x) b_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha' h(\alpha) h(\alpha') d\alpha d\alpha' + o(b_n)$$

$$= b_n^2 (f(x, x) + o(b_n)).$$

(2.61)

Dealing with the second term of $\gamma(t)$

$$E(z_1)^2 = \left(\int_{-\infty}^{\infty} h\left(\frac{x-u}{b_n}\right) f(u) du\right)^2$$

$$= (b_n \int_{-\infty}^{\infty} h(\alpha) f(x - b_n \alpha) d\alpha)^2$$

$$= (b_n f(x) + b_n (o(b_n)))^2$$

(2.62)
therefore

\[ \gamma(t) = b_n^2 (f(x, x) + o(n) - (f(x))^2) \]

\[ \approx o(n) \] (2.63)

thus the variance of the estimated marginal density is asymptotically 0. Now dealing with the bias of the estimate marginal density

\[ Bias = E(\tilde{f}(x)) - f(x) \] (2.64)

where

\[
E(\tilde{f}(x)) = E\left( \frac{1}{nb_n} \sum_{i=1}^{n} h \left( \frac{x - X_i}{b_n} \right) \right) \\
= \frac{1}{b_n} \int_{-\infty}^{\infty} h \left( \frac{x - u}{b_n} \right) f(u) du \\
= \int_{-\infty}^{\infty} h(\alpha) f(x - b_n \alpha) d\alpha \\
= f(x) + o(b_n). \] (2.65)

Therefore the bias is

\[ Bias = f(x) + o(b_n) - f(x) \]

\[ = o(b_n) \] (2.66)

The bias is 0 since both the bias and variance go \( \to 0 \) as \( n \to \infty \) therefore the estimated marginal density is a consistent estimator of the true marginal density. Since the estimate of both the joint and the marginal densities are both consistent estimators thus the estimated conditional density is also consistent estimate of the true conditional density.
2.11 Simulation

First of all, we recall the following important property about conditional density: if $X$ and $Y$ are jointly normally distributed with means $\mu_X$ and $\mu_Y$, variances $\sigma_X^2$ and $\sigma_Y^2$, and correlation coefficient $\rho$, then

$$f_{Y|X}(y|x) = \mathcal{N}(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \sigma_Y^2(1 - \rho^2)).$$

The data of this section come from AR(1) model: $X_i = \phi X_{i-1} + \epsilon_i$, where there are 20 samples each having 100 observations. The data points have a distribution of $X_i \sim N(0, \frac{1}{1-\rho^2})$, where $\rho = \phi$, and we take $Y_i = X_{i+1}$. The conditional distribution of $Y_i$ when $X_i = x$ is

$$N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x - \mu_2), \sigma_1^2(1 - \rho^2)) = N(\rho x, 1) \quad (2.67)$$

where $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. For this section there are 3 different models where the first being a stationary AR(1) model where $\phi = 0.6$ The second model is close to being non-stationary where $\phi = 0.8$, the last model is close to the data being independent and identically distributed (i.i.d.) random variables where $\phi = 0.2$.

### Table 1

Optimal Bandwidths from Integrated Squared Error and Log Likelihood

<table>
<thead>
<tr>
<th>Method</th>
<th>$I(\hat{f})$</th>
<th>$I(\tilde{f})$</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISE(Oracle)</td>
<td>$6.99 \times 10^{-3}$</td>
<td></td>
<td>0.424</td>
<td>0.439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISE(CV)</td>
<td>0.610721</td>
<td>1.20872</td>
<td>0.552</td>
<td>0.62</td>
<td>0.54</td>
<td>1.511</td>
<td>1.628</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>5.84812</td>
<td>1.31912</td>
<td>0.544</td>
<td>0.54</td>
<td>0.545</td>
<td>0.677</td>
<td></td>
</tr>
</tbody>
</table>

$I(\tilde{f})$ is for the one step estimator and parameters $h_1$ and $h_2$ are the bandwidth parameters for the one step estimator. $I(\hat{f})$ is for the 2 step estimator and $b_0, b_1, b_2$ are the bandwidth
parameters for the 2 step estimator. In the 2nd row of table 1 is of the infeasible oracle bandwidth values. All of the places for the 2 step estimator is left blank on the 2nd row, this is due to the fact that double integration of the 2 step estimator took longer than 48 hours to compute. Also to save time the double integration was done between \([-10, 10]\) instead of \((-\infty, \infty)\). Since the mean of the conditional density is \(\rho_y\) which is \(\approx 0\) thus taking any other bound larger than \([-10, 10]\) would probably result in a very minimal change. Also the bounds used for the integrations resulted in values very close to the true density as can be seen in the first value of the 2nd row in table 1 therefore the bandwidth values found for the one step bandwidth are the optimal values of those bandwidths. The values of table 1 and all the other tables are of medians of the 20 samples. The reason that medians are used instead of means is because when finding the optimal bandwidth from the data some of the samples bandwidths were found \(\to -\infty\) and throwing the mean values of the bandwidths \(\to -\infty\). Looking at the bandwidth values of the one step estimator of the 3rd row, they are very close to those of the oracle method which is a very encouraging result and shows how strong this method is. When comparing the ISE score of the one step and two step estimators, the score of the one step is smaller and thus one could conclude that one step estimator seems to be the better estimator when dealing with AR(1) data. The last row of table 1 is for the log likelihood cross validation method. When comparing the Kullback-leibler scores of the two conditional density estimators it can be seen that the two step estimator has much lower value thus the two step estimator is considered to be the better technique of estimating the true density. Comparing the parameter values of the one step estimator of the log likelihood to the MISE method it could be seen that the log likelihood method is closer to the oracle method bandwidth. Therefore the log likelihood method can be concluded to be the better estimator of the true density and also when using the log likelihood method the bandwidth parameters none of the samples bandwidth estimates did explode to \(-\infty\).

Figure 2.1 is of the true density which is coloured black, the red density curve is of the one step log likelihood cross validation method. The green curve is of the two step estimate from
the log likelihood estimation technique. The bandwidths obtained from the ISE oracle and 
CV were not shown since they took so long to obtain.

Figure 2.1:

Since the log likelihood method have very similar curves to the true density and also since 
the log likelihood curve has the quickest computational time this will be the method used to estimate the rest of the densities.
AR(1) with $\phi = 0.8$ and 0.2 conditional densities were also estimated.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
$\phi$ & $I(\hat{f})$ & $I(\hat{f})$ & $h_1$ & $h_2$ & $b_0$ & $b_1$ & $b_2$ \\
\hline
0.8 & 6.06862 & 1.59912 & 0.556 & 0.577 & 1.031 & 0.749 & 1.18 \\
0.2 & 5.67512 & 1.13112 & 0.526 & 0.539 & 3.319 & 0.534 & 0.552 \\
\hline
\end{tabular}
\caption{Optimal Bandwidths Obtained from Log Likelihood}
\end{table}

When comparing the 3 log likelihood cross validation Kullback-leibler scores from tables 1&2 it can be seen that closer you get to iid dataset the lower the Kullback-leibler scores will be. Also the trend for the bandwidth parameters is the same except for the bandwidth of the estimate of the mean, $b_0$ does not seem to follow that tendency. It seems as though the further away it is of being iid and non-stationary the lower the score of that bandwidth will be.

Also comparing figures 2.1, 2.2 and 2.3 it could be seen that the closer you get to non-stationary dataset the weaker the conditional density estimator estimates the true conditional density.

A data set composed of only ARCH(1) errors was also studied. This ARCH(1) data was simulated with $\alpha_0 = 2$ and $\alpha_1 = 0.7$.

From figure 2.5 it could be seen that the data has some volatility. The next step is to look at the auto-correlation function (acf) and partial auto-correlation function (pacf).

From figure 2.7 it seems as though $y_t$ and $x_t$ where $x_t = y_{t-1}$ are independent of each there. When the data is squared and the ACF and PACF are observed, it could be seen from figure 2.7 that the ACF seems to be dying down while the PACF cuts off after lag 1 therefore proving that this data is indeed ARCH(1) model and when squared
becomes a non-Gaussian AR(1) model.

\[ \alpha_0 = 2 \quad \& \quad \alpha_1 = 0.7 \]

<table>
<thead>
<tr>
<th>(I(\hat{f}))</th>
<th>(I(\tilde{f}))</th>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(b_0)</th>
<th>(b_1)</th>
<th>(b_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.21712</td>
<td>7.23112</td>
<td>1.332</td>
<td>1.111</td>
<td>2.323</td>
<td>1.36</td>
<td>1.108</td>
</tr>
</tbody>
</table>

Looking at table 3 \(I(\hat{f})\) has a lower score than \(I(\tilde{f})\) thus the conclusion that the one step estimator is the better estimate of the true conditional density. Figure 2.4 agrees with this conclusion where the black curve is of the true conditional density, the green curve is of the one step estimator log likelihood leave one out cross validation and the red curve is of the 2 step method.

The conclusion can be made that when dealing with ARCH(1) model whether the model is close to i.i.d random variables data or close to non-stationary data that the 2 step log likelihood leave one out cross validation estimator is the better choice out of the two. When the data are not auto-correlated there does not seem to be any difference between the estimation of the conditional density using the one step or two step method. From table 3 it can be seen that the one step estimator estimated the conditional density better than the two step method. These results can even be seen in figure 5.

Finally a data set of an AR(1) process with ARCH(1) error was also studied. Just like the other data sets this data set is composed of 20 samples with each sample having a sample size of 100. The parameter of the AR(1) part is \(\phi = 0.6\) and the parameters for the error are \(\alpha_0 = 2\), \(\alpha_1 = 0.7\). The results of the Kullback-leibler information and optimal bandwidth windows are list in table 4.

\[\text{Table 4}\]

30
Bandwidth Values Obtained From Log Likelihood of AR(1) with ARCH(1) errors

\[ \alpha_0 = 2 \quad & \quad \alpha_1 = 0.7 \quad \phi = 0.6 \]

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( I(\hat{f}) )</td>
<td>( I(\tilde{f}) )</td>
<td>( h_1 )</td>
<td>( h_2 )</td>
<td>( b_0 )</td>
<td>( b_1 )</td>
<td>( b_2 )</td>
</tr>
<tr>
<td>7.427668</td>
<td>7.413447</td>
<td>0.95156</td>
<td>0.99312</td>
<td>1.122402</td>
<td>1.183994</td>
<td>1.100831</td>
</tr>
</tbody>
</table>

The two step process has a lower \( I \) score compared to the one step as can be seen in table 4. Due to this fact, the conclusion could be made that the two step process is better estimate of the true conditional density.
Figure 2.2:
Figure 2.3: AR(1), phi=0.2
Figure 2.4:
Figure 2.5:
Figure 2.6:

ACF and PACF of ARCH(1)
Chapter 3

Multivariate Case

3.1 One Step Estimator

Multivariate conditional density is very similar to the univariate one. For example, the conditional density of $Y$ given $(X_1, X_2)$ is

$$f(y \mid x_1, x_2) = \frac{f(x_1, x_2, y)}{f(x_1, x_2)}.$$  (3.1)

The numerator is the joint density of $(X_1, X_2, Y)$ and the denominator is the joint density of $(X_1, X_2)$. The multivariate density and the univariate density have similar properties. The denominator of the conditional density is estimated via one step estimator by

$$\tilde{f}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} K_h(\|x - X_i\|)$$  (3.2)

where $X = (X_1, X_2)$. The numerator of the conditional density is estimated using the one step estimator by

$$\tilde{f}(x_1, x_2, y) = \frac{1}{n} \sum_{i=1}^{n} K_{h_2}(\|x - X_i\|) K_{h_1}(y - y_i).$$  (3.3)
From (3.2) and (3.3) the one step kernel estimation of the conditional density is calculated

\[
\tilde{f}(y \mid x_1, x_2) = \frac{\tilde{f}(x_1, x_2, y)}{\tilde{f}(x_1, x_2)} = \frac{\sum_{i=1}^{n} K_{h_2}(\|x - X_i\|)K_{h_1}(y - y_i)}{\sum_{i=1}^{n} K_h(\|x - X_i\|)}
\]

(3.4)

where

\[
\|x - X_i\| = \sqrt{(x - X_{1i})^2 + (x - X_{2i})^2}
\]

(3.5)

for \(i = 1, 2, ..., n\). This concept originated from the combination of Hansen (2004) and Holmes et al (2007).

### 3.2 Two Step Estimator

The two step estimator is calculated the same way as in the univariate case where the first step is estimating the conditional mean. The estimated conditional mean for the multivariate case is

\[
\hat{m}(x) = \frac{\sum_{i=1}^{n} K_{b_2}(\|x - X_i\|)Y_i}{\sum_{i=1}^{n} K_{b_2}(\|x - X_i\|)}.
\]

(3.6)

This is just an extension of the univariate case as was shown in Hansen (2004). The second step is estimating the multivariate conditional density and this is accomplished by

\[
\hat{f}(y \mid x) = \frac{\sum_{i=1}^{n} K_{b_2}(\|x - X_i\|)K_{b_1}(y - \hat{m}(x) - \hat{\epsilon}_i)}{\sum_{i=1}^{n} K_{b_2}(\|x - X_i\|)}.
\]

(3.7)
3.3 Bandwidth Selection

Just like the univariate case the bandwidth scores were estimated using the cross validation leave on out log likelihood where \( x = (x_1, x_2) \). The log likelihood function for the one step is defined as

\[
L = \frac{1}{n} \sum_{i=1}^{n} \log(\tilde{f}_{-i}(Y_i \mid X_i)\tilde{f}_{-i}(X_i))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\sum_{j \neq i} K_{h_1}(Y_i - Y_j)K_{h_2}(\|X_i - X_j\|)}{\sum_{j \neq i} K_{h_2}(\|X_i - X_j\|)} \right) + \log \left( \frac{1}{n-1} \sum_{j \neq i} K_{h_2}(\|X_i - X_j\|) \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \log \left( \sum_{j \neq i} K_{h_1}(Y_i - Y_j)K_{h_2}(\|X_i - X_j\|) \right) - \log(n-1).
\] (3.8)

The likelihood cross validation choice of \( h_1 \) and \( h_2 \) are the values of \( h_1 \) and \( h_2 \) which minimizes the score of \( I \). Minimizing the score of \( I \) results in an estimated density that is close to the true density. For the 2 step estimator the optimal bandwidth scores are obtained via 2 steps. The first step is to find the optimal value for the Nadaraya-Watson estimator using the leave one out cross validation method

\[
\hat{m}_{-i}(X_i) = \frac{\sum_{j \neq i} K_{h_0}(\|X_i - X_j\|)Y_j}{\sum_{j \neq i} K_{h_0}(\|X_i - X_j\|)}.
\] (3.9)

Once the optimal bandwidth value for the Nadaraya-Watson estimator is obtained it is used to find the best bandwidth scores for the conditional density estimate. The optimal bandwidth scores are found using the log likelihood

\[
L = \frac{1}{n} \sum_{i=1}^{n} \log(\hat{f}_{-i}(Y_i \mid X_i)\hat{f}_{-i}(X_i))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\sum_{j \neq i} K_{h_1}(Y_i - \hat{m}(X_i) - \hat{e}_i)K_{h_2}(\|X_i - X_j\|)}{\sum_{j \neq i} K_{h_2}(\|X_i - X_j\|)} \right) + \log \left( \frac{1}{n-1} \sum_{j \neq i} K_{h_2}(\|X_i - X_j\|) \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \log \left( \sum_{j \neq i} K_{h_1}(Y_i - \hat{m}(X_i) - \hat{e}_i)K_{h_2}(\|X_i - X_j\|) \right) - \log(n-1).
\] (3.10)
For simulation purposes where the true density is known the leave one out cross validation log likelihood method could be checked to see how good it estimates for the bandwidth values by performing the Kullback-leibler information which is just a multivariate extension of the idea from Silverman (1986)

\[
I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log \left( \frac{f(y \mid x)}{\hat{f}(y \mid x)} \right) dy dx
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log(f(y \mid x))dy dx - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \log(\hat{f}(y \mid x))dy dx
\]  

(3.11)

where \( x = (x_1, x_2) \). Just like the univariate case only the negative term of (3.11) is calculated since only this part contains the bandwidth parameters. In the previous chapter it was shown that density estimation for a joint density is a consistent estimator thus the multivariate conditional density estimation is also a consistent and asymptotically unbiased estimator of the true multivariate conditional density, seeing that a multivariate conditional density is just a ratio of 2 joint densities as was shown in (3.1).

### 3.4 Multivariate Response

A multivariate response case where \( y \) is the response variable is where there are several response variables \( y_1, y_2, ..., y_p \). For this section we’ll be dealing with the simplest case where there are only two response variables and for this section to make the notation easier we’ll let \( y = (y_1, y_2) \) and the independent variables \( x_1, x_2, .., x_r \) will follow the same manner as the previous section where there only 2 independent variables and the vector of \( (x_1, x_2) \) is denoted as \( x \). The conditional density

\[
f(y \mid x) = \frac{f(x_1, x_2, y_1, y_2)}{f(x_1, x_2)}
\]  

(3.12)
is estimated via one step estimator by

\[
\tilde{f}(y \mid x) = \frac{\sum_{i=1}^{n} K_{h_1}(\|x - X_i\|) K_{h_2}(\|y - Y_i\|)}{\sum_{i=1}^{n} K_{h_1}(\|x - X_i\|)}
\]

(3.13)

where again \(Y_i = X_{t+1}\). The optimal bandwidth parameters are calculated by leave one out log likelihood cross validation

\[
L = \frac{1}{n} \sum_{i=1}^{n} \log \left( \sum_{j \neq i} K_{h_2}(\|Y_i - Y_j\|) K_{h_1}(\|X_i - X_j\|) \right) - \log(n - 1).
\]

(3.14)

The above equation is then optimized to obtain argmin of \(L\) which result in the optimal bandwidth scores. For the two step estimator the first step is to start with \(Y_t = m(X_t) + \epsilon_t\) where \(Y_t, X_t\) and \(m(X_t)\) are all vectors. Since \(m(X_t)\) and \(\epsilon_t\) are unknown, hence \(m(x)\) is estimated as

\[
\hat{m}(x) = \frac{\sum_{i=1}^{n} K_{b_0}(\|x - X_i\|) Y_i}{\sum_{i=1}^{n} K_{b_0}(\|x - X_i\|)}.
\]

(3.15)

The leave one out cross validation technique is applied again to get the best score for the smoothing operator \(b_0\) which is calculated by

\[
\hat{m}(X_i) = \frac{\sum_{j \neq i} K_{b_0}(\|X_i - X_j\|) Y_j}{\sum_{j \neq i} K_{b_0}(\|X_i - X_j\|)}
\]

(3.16)

where \(\hat{m}(X_i)\) is a vector made up of \((\hat{m}_1(X_{1i}), \hat{m}_2(X_{2i}))\). This is followed up by estimating the residuals \(\epsilon_t\) as

\[
\hat{\epsilon}_t = Y_t - \hat{m}(X_t)
\]

(3.17)
where \( \hat{\epsilon}_t \) is a vector made up as \( \hat{\epsilon}_t = (\hat{\epsilon}_{1t}, \hat{\epsilon}_{2t}) \). The vector \( \hat{\epsilon} \) is then used in the conditional density estimation of the two step estimator as

\[
\hat{f}(y \mid x) = \frac{\sum_{i=1}^{n} K_{b_2} (\|x - X_i\|) K_{b_1} (\|y - \hat{m}(x) - \hat{\epsilon}_i\|)}{\sum_{i=1}^{n} K_{b_2} (\|x - X_i\|)}.
\]

(3.18)

The leave one out cross validation log likelihood method is employed to acquire the ideal bandwidth scores,

\[
L = \frac{1}{n} \sum_i \log(f_{-i}(Y_i \mid X_i)\hat{f}_{-i}(X_i))
\]

\[
= \frac{1}{n} \sum_i \log \left( \frac{\sum_{j \neq i} K_{b_1} (\|e_i - e_j\|) K_{b_2} (\|X_i - X_j\|)}{\sum_{j \neq i} K_{b_2} (\|X_i - X_j\|)} \right) + \log \left( \frac{1}{n-1} \sum_{j \neq i} K_{b_2} (\|X_i - X_j\|) \right)
\]

\[
= \frac{1}{n} \sum_i \log(\sum_{j \neq i} K_{b_1} (\|e_i - e_j\|) K_{b_2} (\|X_i - X_j\|)) - \log(n-1)
\]

(3.19)

since \( \log(n-1) \) does not contain a bandwidth parameter therefore it could be disregarded.

The Kullback-leibler information is calculated exactly the same way as the previous section also the previous chapter already proves that the estimated conditional density is a consistent and asymptotically unbiased estimator.

### 3.5 Simulation

The data of this section come from \( AR(2) \) model, where there are 20 samples each having 100 observations. In an \( AR(2) \) \( X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \) where \( \delta \) is a constant and just like the univariate case \( \epsilon \) has a \( N(0, \sigma^2) \) distribution (see Shumway (2011)). Here \( X_t = (X_{t-1}, X_{t-2}) \) and \( Y_t = X_t \). \( AR(2) \) is stationary if \( \phi_2 \pm \phi_1 < 1 \). By stationarity, the mean of \( AR(2) \) process
can be calculated as follows

\[ X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \]

\[ E(X_t) = E(\delta) + E(\phi_1 X_{t-1}) + E(\phi_2 X_{t-1}) \]

\[ \mu = \delta + \phi_1 \mu + \phi_2 \mu \]

\[ = \frac{\delta}{1 - \phi_1 - \phi_2} \]  \hspace{1cm} (3.20)

and using (3.20) the variance and autocovariance functions are calculated

\[ X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \]

\[ = \mu(1 - \phi_1 - \phi_2) + \phi_1 X_{t-1} + \phi_2 X_{t-2} \]

\[ X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \epsilon_t \]

\[ E(X_t - \mu)(X_{t-i} - \mu) = \phi_1 E(X_{t-1} - \mu)(X_{t-i} - \mu) + \phi_2 E(X_{t-2} - \mu)(X_{t-i} - \mu) + E(\epsilon_t(X_{t-i} - \mu)) \]

\[ \gamma_i = \phi_1 \gamma_{i-1} + \phi_2 \gamma_{i-2}. \]  \hspace{1cm} (3.21)

The variance of the AR(2) process is calculated when \( i = 0 \), the autocovariance function at lag 1 is obtained by substituting 1 into where \( i \) is in the equation (3.21) and the autocovariance
function is calculated when $i = 2$.

\[
\begin{align*}
\gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2 \epsilon \\
\gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 \\
\gamma_1(1 - \phi_2) &= \phi_1 \gamma_0 \\
\gamma_1 &= \frac{\phi_1}{1 - \phi_2} \gamma_0 \\
\gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 \\
&= \frac{\phi_1^2}{1 - \phi_2} \gamma_0 + \phi_2 \gamma_0 \\
&= \frac{\phi_1^2 + \phi_2(1 - \phi_2)}{1 - \phi_2} \gamma_0 \\
\gamma_0 &= \frac{\phi_1}{1 - \phi_2} \gamma_0 + \phi_2(\phi_1^2 + \phi_2(1 - \phi_2)) \gamma_0 + \sigma^2 \epsilon \\
&= \frac{\phi_1^2 + \phi_2 \phi_1^2 + \phi_2^2(1 - \phi_2)}{1 - \phi_2} \gamma_0 + \sigma^2 \epsilon \\
\sigma^2 \epsilon &= \frac{1 - \phi_2 - \phi_1^2(1 + \phi_2) - \phi_2^2(1 - \phi_2)}{1 - \phi_2} \\
&= \frac{\gamma_0 (1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)}{1 - \phi_2} \\
\gamma_0 &= \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)} \sigma^2 \epsilon \\
&= \frac{\phi_1 \sigma^2 \epsilon}{(1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)} \\
\end{align*}
\]

thus $AR(2)$ process $\sim N\left(\frac{\delta}{1-\phi_1-\phi_2}, \frac{1-\phi_2}{(1+\phi_2)(1-\phi_2-\phi_1)(1-\phi_2+\phi_1)} \sigma^2 \epsilon \right)$ (see Brockwell and Davis (1991))

the autocovariance function at lag 1 is

\[
\begin{align*}
\gamma_1 &= \frac{\phi_1}{1 - \phi_2} \gamma_0 \\
&= \frac{\phi_1 \sigma^2 \epsilon}{(1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)} \\
&= \frac{\phi_1 \sigma^2 \epsilon}{(1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)} \quad (3.23)
\end{align*}
\]
and at lag 2
\[
\gamma_2 = \frac{\phi_1^2 + \phi_2(1 - \phi_2)}{1 - \phi_2} \gamma_0 = \frac{\phi_1^2 + \phi_2(1 - \phi_2)}{(1 + \phi_2)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)} \sigma^2, \tag{3.24}
\]

With this information a bivariate normal distribution which is \(\sim N_2(\mu_1 - \Sigma_{12}^{-1}\mu_2 + \Sigma_{12}^{-1}X_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, 1)\), where \(\Sigma_{11} = \gamma_0, \Sigma_{12} = (\gamma_1 \gamma_2)\) and \(\Sigma_{21}\) is the transpose of \(\Sigma_{12}\). Finally \(\Sigma_{22}\) is a two by two matrix defined as
\[
\Sigma_{22} = \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{pmatrix} \tag{3.25}
\]

There were 3 different models tested with all having 20 samples each and each sample having 100 data points and \(\delta = 0\) and \(\sigma^2 = 1\) for all the models. This leads to the bivariate normal distribution \(N(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\). The first model for this section was the stationary case where \(\phi_1 = 0.3, \phi_2 = 0.5\) thus the AR(2) has a distribution \(\sim N(0, 2.08333)\) and the conditional distribution is \(N(-0.4397114, 1)\). The results of this sample are found in table 5.

**Table 5**

Log Likelihood Scores and Optimized Bandwidth of Stationary Case

\(\phi_1 = 0.3, \phi_2 = 0.5\)

<table>
<thead>
<tr>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(I(\hat{f}))</th>
<th>(b_0)</th>
<th>(b_1)</th>
<th>(b_2)</th>
<th>(I(\tilde{f}))</th>
<th>(I(\tilde{f}) - I(\hat{f}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45318145</td>
<td>1.68038905</td>
<td>6.548204</td>
<td>1.677739822</td>
<td>0.48477</td>
<td>1.887129</td>
<td>6.473862</td>
<td>0.074342</td>
</tr>
</tbody>
</table>

The values of table 5 are the medians of the 20 samples. The last column in table 5 is calculated by \(I(\tilde{f}) - I(\hat{f})\). From that column of table 5 it could be seen that estimating \(f(y \mid x)\) via the two step has a lower log likelihood value thus it is closer to the true density.
Also comparing $h_2$ and $b_2$ it could be seen that $b_2 > h_2$ meaning that there is less smoothing done on the two step estimator method compared to the one step estimation method of the conditional density.

From figure 3.1 it could be seen that both these methods are good estimators for the true conditional density. The green curve is of the two-step process while the red curve is of the one step method.

AR(2) model with $\phi_1 = 0.2$, $\phi_2 = 0.15$ therefore this data is from a normal distribution with mean of 0 and variance of 1.082975 while the conditional distribution is $N(0.2215732, 1)$. 

47
Table 6 lists the results of this simulation

\[ \phi_1 = 0.2, \phi_2 = 0.15 \]

<table>
<thead>
<tr>
<th>( h_1 )</th>
<th>( h_2 )</th>
<th>( I(\tilde{f}) )</th>
<th>( b_0 )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( I(\hat{f}) )</th>
<th>( I(\tilde{f}) - I(\hat{f}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.44805655</td>
<td>1.5424916</td>
<td>6.167818</td>
<td>0.140669681</td>
<td>0.3545089</td>
<td>1.56923105</td>
<td>5.945063</td>
<td>0.222755</td>
</tr>
</tbody>
</table>

Just like table 5 the last column displays the subtraction of the two log likelihood scores since it is positive thus the two step estimator is better of the true conditional density.

From figure 3.2 it could be seen that the one step estimator is much better than the two step. It could be said that when the data is close to the i.i.d. case that the two-step estimator is no good at estimating the true conditional density.

AR(2) model with \( \phi_1 = 0.5, \phi_2 = 0.45 \) with a distribution of \( N(0, 7.224959) \) and the conditional distribution is \( N(-0.2812224, 1) \). For the AR(2) case the process is said to be stationary if \( \phi_2 \pm \phi_1 < 1 \). This data set is close to being non-stationary since \( \phi_1 + \phi_2 = 0.95 \). Table 7 lists the results of this simulation

\[ \phi_1 = 0.5, \phi_2 = 0.45 \]

<table>
<thead>
<tr>
<th>( h_1 )</th>
<th>( h_2 )</th>
<th>( I(\tilde{f}) )</th>
<th>( b_0 )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( I(\hat{f}) )</th>
<th>( I(\tilde{f}) - I(\hat{f}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.464837</td>
<td>1.715683</td>
<td>6.950714</td>
<td>2.014217</td>
<td>2.030661</td>
<td>5.721551</td>
<td>1.229163</td>
<td></td>
</tr>
</tbody>
</table>

Just like tables 5 and 6 the log likelihood scores of the 2-step conditional estimation process are smaller than those of the one step. In fact on average the 2-step process log likelihood
scores were almost half the values when compared to those of the one-step process. From table 7 it could be seen that $b_2 > h_2$ meaning that less smoothing is done in the 2-step process than the 1-step process as is explained in Hansen (2004).

When the data is close to the non-stationary case the curves are very rigid and lose their smoothness.

When comparing the three different models the last model had the biggest difference in $I(\hat{f})$ and $I(\hat{f})$ is the model with $\phi_1 = 0.5, \phi_2 = 0.45$. Comparing figures 3.1, 3.2, and 3.3, it
can be seen that the biggest difference of the one and two step processes is in figure 3.2.
Chapter 4

Multivariate Case With Real World Applications

4.1 Stock Market Data

The first data set of this chapter is of stock market daily returns of the New York Stock Exchange (NYSE) and Toronto Stock Exchange (TSE) from June 9th 2011 to June 9th 2014. The conditional densities were estimated as if they were AR(1) type where tomorrow’s forecast price is conditioned on one day prior’s price. Both the one and two step processes are done on both stock prices. Also the density forecast is built from the TSX and NYSE separately conditioned on both stock trade prices. The fourth method is building a joint density estimate of NYSE and TSX conditioned on the previous day’s price of both the NYSE and TSX then the marginal density is obtained by integrating out the unwanted stock market. Looking at the TSX first, the first method is the one step process where the conditional forecast density is \( f(y_2 | x_2) \) where \( y_2 \) is the price of tomorrow, \( x_2 \) is the stock market price of TSX today, \( x_1 \) is the price of today’s stock price of NYSE and \( y_1 \) is the tomorrow’s NYSE stock prices. The
results of the optimal bandwidth windows and log likelihood scores can be seen in table 8.

Table 8

<table>
<thead>
<tr>
<th>TSX One Step Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_1$</td>
</tr>
<tr>
<td>55.56</td>
</tr>
</tbody>
</table>

The bandwidth windows are much larger than the previous chapters and this is due to the fact the stock market prices of the TSX are so large. These prices range from 11177.91 to 14800.18 which are values that are much larger than the previous.

Table 9 is of the two step univariate process

Table 9

<table>
<thead>
<tr>
<th>TSX Two Step Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
</tr>
<tr>
<td>57.48762</td>
</tr>
</tbody>
</table>

Comparing $b_2$ to $h_2$ it could be seen that $b_2 > h_2$ thus the conclusion can be made that less smoothing is required for the two step. Also the $b_2$ value is much larger than that of the $b_1$ score and as was stated in Hansen (2004) the value of $b_2$ means that cross-validation eliminates the conditional smoothing in the second step, so that the estimated conditional density only depends on $x_t$ through the estimated conditional mean.

Table 10

$f(y_2 | x_1, x_2)$
When comparing the log likelihood scores from tables 8, 10, 11 it could be seen that the lowest score is from the method of table 8 for that reason the two step process is only done for that method.

The black curve is of the forecast density of the marginal from the joint conditional density. The red curve is of the one step process, green is of the two step process and the blue curve is of the forecast density of the tomorrow’s price conditioned on the joint density of today’s TSX and NYSE prices. Looking at curves of figure 4.1 and also looking at the data from the tables in this section it could be concluded that the marginal forecast density from the joint forecast density is the weakest method of estimating the true density.

| One Step | Two Step | \( f(y_2 | x_1, x_2) \) | marginal from joint |
|----------|----------|----------------|-------------------|
| 0.07486203 | 0.08002461 | 0.0980361 | 0.03998905 |

The values of table 12 correspond to

\[
\int_{x_{750-10}}^{x_{750+10}} \text{density forecast}(y)dy
\]
where density forecast is one of the four density forecasts ($\hat{f}$, $\tilde{f}$, conditional multivariate, or the marginal conditional multivariate). All these probability values are all close to each other but as expected the marginal forecast density obtained from the joint density is the weakest method of forecasting tomorrow’s price. The best method seems to be the method that takes advantage of the dependence of the TSX prices on the NYSE. That is unexpected since the log likelihood score of the two step method is much lower than that of this method. There are a few possible methods why the forecast density of the marginal conditional density from the joint conditional density is so poor. The first reason could be the fact that when integrating out the $y_1$ variable, the joint conditional density forecast could open the door to the possibilities of errors occurring. Another possible method would be to integrate out $y_2$ first and then optimize for the bandwidth scores. The first step would be build the joint conditional density

$$
\hat{f}(y_1, y_2 \mid x_1, x_2) = \frac{\sum_i K_h(\|y - Y_i\|)K_h(\|x - X_i\|)}{\sum_i K_h(\|x - X_i\|)}.
$$

(4.1)

The next step is to integrate with respect to $y_2$

$$
\hat{g}(y_1 \mid x_1, x_2) = \int_{-\infty}^{\infty} \hat{f}(y \mid x)dy_2
$$

(4.2)

where $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Referring back to Holmes et al (2007) for the calculation of the log likelihood

$$
L = \frac{1}{n} \sum_i \log(\hat{g}(y_1 \mid x_1, x_2))\hat{g}(x)
$$

(4.3)

where $\hat{g}$ is defined as

$$
\hat{g}(x) = \int_{-\infty}^{\infty} \hat{f}(y_2, x)dy_2.
$$

(4.4)
Continuing on from (4.3) leads to the equation of the leave one out log likelihood equation for this method as

\[
L = \frac{1}{n} \sum_i \log(\int_{-\infty}^{\infty} A_i dy_2) + \log\left(\frac{1}{n-1} \int_{-\infty}^{\infty} C_i dy_2\right)
\]

\[
= \frac{1}{n} \sum_i \log(\int_{-\infty}^{\infty} A_i dy_2) - \log(B_i) + \log(\int_{-\infty}^{\infty} C_i dy_2) - \log(n-1)
\]

(4.5)

where \( A_i = \sum_{j \neq i} K_{h_1}(\|Y_i - Y_j\|)K_{h_2}(\|X_i - X_j\|) \), \( B_i = \sum_{j \neq i} K_{h_2}(\|X_i - X_j\|) \) and \( C_i = \sum_{j \neq i} K_{h_1}(Y_i - Y_j)K_{h_2}(\|X_i - X_j\|) \). This method seems more logical than the previous procedure. This method seems reasonable and since it takes advantage of the dependence of the TSX on the NYSE. The computational part of this method was not included but will be part of future work.

The same procedure was done on the NYSE data set.

**Table 13**

<table>
<thead>
<tr>
<th>NYSE One Step Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_1 )</td>
</tr>
<tr>
<td>55.24414</td>
</tr>
</tbody>
</table>

The bandwidth windows are a bit smaller than those of the TSX and this is due to the fact the stock market prices of the NYSE are lower than those of those TSX. These prices range from 6574.29 to 10904.22 where the maximum price value of NYSE is less than that of minimum of TSX.

Table 14 is of the two step process

**Table 14**

<table>
<thead>
<tr>
<th>NYSE Two Step Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 55 )</td>
</tr>
</tbody>
</table>
Comparing $b_2$ to $h_2$ it could be seen that $b_2 > h_2$ thus the conclusion can be made that less smoothing is required for the two step just like in the case of TSX data. Also the $b_2$ value is much larger than that of the $b_1$ score and as was stated in Hansen (2004) the value of $b_2$ means that cross-validation eliminates the conditional smoothing in the second step, so that the estimated conditional density only depends on $x_t$ through the estimated conditional mean.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
$b_0$ & $b_1$ & $b_2$ & $I(\hat{f})$ \\
\hline
49.27849 & 39.93191 & 140.16763 & 12.19342 \\
\hline
\end{tabular}
\caption{Table 15}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
$h_1$ & $h_2$ & $I(\hat{f})$ \\
\hline
59.90662 & 58.93257 & 18.07153 \\
\hline
\end{tabular}
\caption{Table 16}
\end{table}

Again the weakest method seems to be forecast density that obtained from the joint conditional density. This is probably due to the facts that were already stated when describing the TSX data.

The black curve is of the forecast density of the marginal from the joint conditional density. The red curve is of the one step process, green is of the two step process and the blue curve is of the forecast density of the tomorrow’s price conditioned on the joint density of today’s TSX and NYSE prices. Looking at curves of figure 4.2 and also looking at the data from the tables in this section it could be concluded that the marginal forecast density from the joint forecast density is the weakest method of estimating the true density again and is already explained in the TSX section.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\emph{Prediction of Tomorrow's Price} \\
\hline
\end{tabular}
\caption{Table 16}
\end{table}
All these probability values are all close to each other but as expected the marginal forecast density obtained from the joint density is the weakest method of forecasting tomorrow’s price. The best method seems to be the method that takes advantage of the dependence of the NYSE prices on the TSX. That is unexpected since the log likelihood score of the two step method is much lower than that of this method, but the difference between this method and the two step is much smaller when compared to the TSX data. The reason for this is mainly due to the fact that TSX prices depend more on NYSE stock prices more than the vice versa.

### 4.2 Weather Data

One year worth of daily data was collected from the Ottawa region dating from January 1st, 2010 till December 31st, 2010. The categories for the weather data were wind speed in kilometers per hour, pressure in kilo pascal and average temperature in degrees Celsius. Just like the previous section for each category four different forecast density were built. The first being one step process, the second being the two step process, the third being the category that the forecast for time $t$ is based on all the variables on time $t-1$ (i.e. conditioning on the variable of interest and the two other ones as well) and the fourth and final being a joint conditional density being built and calculating for the desired marginal conditional density. For this section the variables with $y$ are the one day ahead while the current day value is represented as $x$. Wind speed is symbolized as $x_1$, $y_1$ while the temperature variable is represented as $x_2$, $y_2$ and pressure variable is symbolized with $x_3$, $y_3$.

The first variable being studied is the daily average temperature. The one and two step process are estimating the conditional density of $f(y_2 \mid x_2)$. The third method is estimating the conditional density $f(y_2 \mid x)$ and the fourth method is estimating the conditional joint density $f(y \mid x)$ and proceeding to accomplish a double integration on $y_1, y_3$, where $y$ is a
vector containing \((y_1, y_2, y_3)\) and \(x\) is also a vector which contains \((x_1, x_2, x_3)\).

Table 17

*One Step procedure for Temperature*

<table>
<thead>
<tr>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(I(\hat{f}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.741543</td>
<td>1.657396</td>
<td>-1.344064</td>
</tr>
</tbody>
</table>

Table 18

*Temperature Two Step Process*

<table>
<thead>
<tr>
<th>(b_0)</th>
<th>(b_1)</th>
<th>(b_2)</th>
<th>(I(\hat{f}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.114106</td>
<td>5.215444</td>
<td>2.269433</td>
<td>-0.7038377</td>
</tr>
</tbody>
</table>

Table 19

*f(y_2 | x)*

<table>
<thead>
<tr>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(I(\hat{f}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6503003</td>
<td>0.9459863</td>
<td>6.73288</td>
</tr>
</tbody>
</table>

Table 20

*Marginal conditional density*

<table>
<thead>
<tr>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(I(\hat{f}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.247444</td>
<td>1.655643</td>
<td>5.711223</td>
</tr>
</tbody>
</table>

When comparing the log likelihood values for the four different methods it could be seen that the one step method is the best at estimating the true density and the extra step of the two step process seems to be unnecessary. Another alarming result is when comparing the \(I\)
values of the tables 19 and 20. The log likelihood score obtained from the joint is much better than that of the third process. The conclusion could be made that the temperature variable does not really depend on the variables. The red curve is of the one step, green is of the two step, blue curve is of the data from table 19 and the black curve is of the data of table 20. Out of the four curves the two step seems to look like the worst and the previous conclusion made seems the worst.

The second variable being studied is the daily average wind speed.

\textit{Table 21}

\textit{One Step procedure for Windspeed}

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
$h_1$ & $h_2$ & $I(\hat{f})$ \\
\hline
1.241889 & 1.513570 & -2.163988 \\
\hline
\end{tabular}
\end{center}

\textit{Table 22}

\textit{Windspeed Two Step Process}

\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
$b_0$ & $b_1$ & $b_2$ & $I(\hat{f})$ \\
\hline
1.966275 & 6.5438612 & 0.7867041 & -1.4764 \\
\hline
\end{tabular}
\end{center}

\textit{Table 23}

\textit{f(y1 | x)}

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
$h_1$ & $h_2$ & $I(\hat{f})$ \\
\hline
3.1488645 & 0.9981343 & 6.978591 \\
\hline
\end{tabular}
\end{center}

Wind speed shows the same trend as temperature that the one step is the best method of estimating the true conditional density.
The colour pattern of figure 4.4 is the same as figure 4.3. From figure 4.4, it could be seen that the two step estimator again is very different from the other three curves and the fact that it has the largest $I$ value which again shows how weak the two-step estimator is. The other three curves are close to each other but again comparing the $I$ values the one step estimator seems to be the best method of estimating the true conditional density of this dataset.

The last variable is pressure the results of this variable are in the proceeding three tables.

\textit{Table 24}

\textit{One Step procedure for Pressure}

<table>
<thead>
<tr>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$I(\hat{f})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2646622</td>
<td>0.1569044</td>
<td>-5.726352</td>
</tr>
</tbody>
</table>

\textit{Table 25}

\textit{Pressure Two Step Process}

<table>
<thead>
<tr>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$I(\hat{f})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1126332</td>
<td>0.9849608</td>
<td>0.2325772</td>
<td>-5.044283</td>
</tr>
</tbody>
</table>

\textit{Table 26}

\textit{$f(y_3 \mid x)$}

<table>
<thead>
<tr>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$I(\hat{f})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5018208</td>
<td>0.5018208</td>
<td>5.189042</td>
</tr>
</tbody>
</table>

The pressure data follows the trends as the other two weather variables where the one step has the smallest $I$ score. These bandwidth windows are the smallest this is due to the fact that the pressure variable has the smallest variance out of the three variables.
Looking at figure 4.5 it seems as though the estimated marginal forecast density from the forecast joint conditional density seems to be the worst and looking at tables 23, 24, 25 and 19 the highest score of $I$ comes from the density where marginal is calculated from the joint which corresponds to the result seen in figure 4.5. The curve of the two step again looks to be a weak curve even though it has a low $I$ score.

Overall from the weather data the one step procedure is concluded to be the best at building the most accurate out of the four methods building forecast conditional density estimate.
Figure 4.2: Conditional Density Estimation of NYSE
Figure 4.3:

Conditional Density Estimations of Temperature
Figure 4.4:

Conditional Density Estimations of Wind Speed
Figure 4.5: Conditional Density Estimations of Pressure
Bibliography


[2] Churvich.pages.stern.nyu.edu/ churvich/Forecasting/Handouts/ARCH.pdf


Chapter 5

Appendix

\begin{verbatim}
AR1.2 <- matrix(NA, nrow=100, ncol=20)

for (i in 1:20){
  AR1.2[, i] <- arima.sim(list(order=c(1,0,0), ar=s), n=100)
}

write.csv(AR1.2, "C:/Users/aziz/Desktop/TimeSeries/AR1.2.csv")

AR1.2 <- read.csv("C:/Users/aziz/Desktop/TimeSeries/AR1.2.csv")

AR1.2 <- AR1.2[, -1]

l1 <- c(1:99); n <- 100

y1.2 <- function(i){AR1.2[2:100, i]}

x1.2 <- function(i){AR1.2[1:99, i]}

kh <- function(x, h){(1/h)*exp(-0.5*(x^2/h^2))}

khx1.2 <- function(l, j, i, h){kh(x1.2[i][l]-x1.2[i][j], h)}

khy1.2 <- function(l, j, i, h){kh(y1.2[i][l]-y1.2[i][j], h)}

Newr1.2 <- function(l, j, i, h){khx1.2(l, j, i, h[2])*khy1.2(l, j, i, h[1])*(1/(n-1))}
\end{verbatim}
Newr11.2 <- function(l, i, h) { log ((sum(Newr1.2(l, l1, i, h))) - Newr1.2(l, l, i, h)) }

Newr21.2 <- function(l, i, h) { (1/n)*Reduce("+", lapply(l1, Newr11.2, i=i, h=h)) }

Newr31.2 <- function(i, h) { -Newr21.2(l, i, h) }

optim(c(1, 1), Newr31.2, control=list(trace=1, maxit=100), i=1)

kb1.2 <- function(x, b) { (1/b)*exp(-0.5*(x^2/b^2)) }

kbo1.2 <- function(l, j, i, b) { kb(x1.2(i)[l] - x1.2(i)[j], b)*y1.2(i)[j] }

kboden1.2 <- function(l, j, i, b) { kb1.2(x1.2(i)[l] - x1.2(i)[j], b) }

mhat1.2 <- function(l, i, b) { (sum(kboden1.2(l, l1, i, b)) - y1.2(i)[1]/b) / (sum(kboden1.2(l, l1, i, b)) - 1/b) }

ehat1.2 <- function(l, i, b) { y1.2(i)[1] - mhat1.2(l, i, b) }

ehatal.2 <- function(l, i, b) { (ehat1.2(l, i, b))^-2 }

lse1.2 <- function(l, i, b) { (1/n)*Reduce("+", lapply(c(1:99), ehatal.2, i=i, b=b)) }

optim(1, lse1.2, method="Brent", lower=0.05, upper=5, i=7)

NN <- NULL

for (m in 1:20) {
  NN[m] <- optim(1, lse1.2, method="Brent", lower=0.05, upper=10, i=m)$par
}

write.csv(optim, "C:/Users/aziz/Desktop/Time Series/NN.csv")

OPT1.2 <- as.matrix(read.csv("C:/Users/aziz/Desktop/Time Series/AR1.2 values.csv"))

h11.2 <- as.numeric(OPT1.2[,1])

h21.2 <- as.numeric(OPT1.2[,2])

b01.2 <- as.numeric(OPT1.2[,4])

wl.2 <- function(l, i) { ehat1.2(l, i, b01.2[i]) }

kkb1.2 <- function(l, j, i, b) { kb(wl.2(1, i) - wl.2(j, i), b) }

Bewr1.2 <- function(l, j, i, b) { khx1.2(l, j, i, b[2]) * kkb1.2(l, j, i, b[1]) }

Bewr11.2 <- function(l, i, h) { log ((sum(Bewr1.2(l, l1, i, h))) - Bewr1.2(l, l, i, h)) }

70
Bewr21.2<-function(i,h){(1/n)*Reduce(”+”,lapply(l1,Bewr11.2,i=i,h=h))}
Bewr31.2<-function(i,h){-Bewr21.2(i,h)}
optim(c(1,1),Bewr31.2,control=list(trace=1,maxit=100),i=1)
OPT1.2<-as.matrix(read.csv("C:/Users/aziz/Desktop/Time_Series/AR1.2 values.csv"))
h1<-as.numeric(OPT1.2[,1])
h2<-as.numeric(OPT1.2[,2])
b0<-as.numeric(OPT1.2[,4])
b1<-as.numeric(OPT1.2[,5])
b2<-as.numeric(OPT1.2[,6])
fx<-function(x,i){dnorm(x,0.2*x1.2(i)[99],1)}
curve(fx(x,1),-10,10)
fdelta<-function(y,i){mconst*(sum(kh(x1.2(i)[99]-x1.2(i),h2[i])*kh(y-y1.2(i),h1[i])))/sum(kh(x1.2(i)[99]-x1.2(i),h2[i]))}
fd<-Vectorize(function(y,i){fdelta(y,i)})
curve(fx(x,1),-10,10)
curve(fd(x,1),-10,10,col=3,add=T)
mm2hat<-function(i){mhat2x(x1.2(i)[99],i,b0[i])}
ehat<-function(1,i,b){y1.2(i)[1]-mhat1.2(1,i,b)}
e1<-function(1,i,b){ehat(1,i,b0[i])}
f1deltanum<-function(s,i,l){mconst*(sum(kh(s-mm2hat(i)-e1(1,i),b1[i])*kh(x1.2(i)[99]-x1.2(i)[1],b2[i])))}
f1deltan<-function(s,i){Reduce(”+”,lapply(l1,f1deltanum,s=s,i=i))}
f1deltad<-function(i){sum(kh(x1.2(i)[99]-x1.2(i),b2[i]))}
f1delta <- function(s, i) {f1deltan(s, i)/f1deltad(i)}
fd1 <- Vectorize(function(y, i) {f1delta(y, i)})
curve(fx(x, 3), -10, 10)
curve(fd(x, 3), -10, 10, col = 3, add = T)
curve(fd1(x, 3), -10, 10, col = 2, add = T)

# where s = 0.6, 0.2 and 0.8

require(fGarch)
ARCH1 <- matrix(NA, nrow = 3000, ncol = 20)
for (i in 1:20)
{
  ARCH111 <- garchSpec(model = list(ar = 0.0, alpha = 0.7, beta = 0, omega = 2))
  ARCH1[, i] <- garchSim(ARCH111, n = 3000)$garch
}l1 <- c(1:2999); n <- 3000
ARCH1 <- read.csv("C:/Users/aziz/Desktop/Time_Series/ARCH1.csv")
ARCH1 <- ARCH1[, -1]
ts.plot(ARCH1[, 1])
require(astsa)
acf2(ARCH1[, 1])
acf2((ARCH1[, 1])^2)
y <- function(i) {ARCH1[2:3000, i]}x <- function(i) {ARCH1[1:2999, i]}
kh <- function(x, h) {(1/h)*exp(-0.5*(x^2/h^2))}
Newr <- function(l, j, i, h) {kh(y(i)[l] - y(i)[j], h[1]) * 
  kh(x(i)[l] - x(i)[j], h[2])/(n - 1)}
Newr1 <- function(l, i, h) {log(sum(Newr(l, 11, i, h)) - Newr(1, l, i, h))}
Newr2 <- function(i, h) {(1/n)*Reduce("+", lapply(l1, Newr1, i = i, h = h))}
Newr3 <- function(i, h) { -Newr2(i, h) }

optim(c(1, 1), Newr3, control=list(trace=1, maxit=300), i=1)

kb <- function(u, b){(1/b)*exp(-0.5*(u^2/b^2))}

kbo <- function(l, j, i, b){kb(x(i)[l]-x(i)[j], b)*y(i)[j]}

kbonum <- function(l, i, b){Reduce("+", lapply(l1, kbo, l=l, i=i, b=b)) - kbo(l, l, i, b)}

kboden <- function(l, j, i, b){kb(x(i)[l]-x(i)[j], b)}

kboden1 <- function(l, i, b){Reduce("+", lapply(l1, kboden, l=l, i=i, b=b)) - kboden(l, l, i, b)}

mhatcv <- function(l, i, b){kbonum(l, i, b)/kboden1(l, i, b)}

ehat <- function(l, i, b){y(i)[l] - mhatcv(l, i, b)}

ehata <- function(l, i, b){(ehat(l, i, b))^2}

lse <- function(i, b){(1/n)*sum(sapply(c(1:99), ehata, i=i, b=b))}

NN <- NULL

for (m in 1:20){

NN[m] <- optim(1, lse, method="Brent", lower=0.00005, upper=5, i=m)$par}

write.csv(NN, "C:/Users/aziz/Desktop/Time_Series/NN.csv")

wl <- function(m, i){ehat(m, i, b0[i])}

kbb1 <- function(l, j, i, b){kb(x(i)[l]-x(i)[j], b)}

kbb2 <- function(l, j, i, b){kb(w1(l, i)-w1(j, i), b)}

Bewr <- function(l, j, i, b){kbb1(l, j, i, b[2])*kbb2(l, j, i, b[1])*1/(n-1)}

Bewr1 <- function(l, i, h){log((sum(Bewr(l, 11, i, h)) - Bewr(l, 1, i, h)))}

Bewr2 <- function(i, h){(1/n)*Reduce("+", lapply(l1, Bewr1, i=i, h=h))}

Bewr3 <- function(i, h){-Bewr2(i, h)}

optim(c(1, 1), Bewr3, control=list(trace=1, maxit=300), i=1)

OPT <- as.matrix(read.csv("C:/Users/aziz/Desktop/Time_Series/ARCH11.csv"))

h1 <- OPT[, 1]
h2<-OPT[,2]
b0<-OPT[,4]
b1<-OPT[,5]
b2<-OPT[,6]
mconst<-1/sqrt(2*pi)

fdelta<-function(y,i){
mconst*(sum(kh(x[i][99]-x(i),h2[i])*kh(y-y(i),h1[i])))
} /sum(kh(x[i][99]-x(i),h2[i]))

fd<-Vectorize(function(y,i){fdelta(y,i)})
mhatn<-function(u,i,h){sum(kh(u-x(i),h)*y(i))}
mhatd<-function(u,i,h){sum(kh(u-x(i),h))}
mhatx<-function(u,i,h){mhatn(u,i,h)/mhatd(u,i,h)}

mmhatx<-function(i){mhatn(x[i][99],i,b0[i])}
e11<-function(1,i){ehat(1,i,b0[i])}

f1deltanum<-function(s,i,l){
mconst*(kh(s-mmhatx(i)-e11(1,i),b1[i])*kh(x[i][99]-x(i)[l],b2[i]))
}

f1deltan<function(s,i){Reduce("+",lapply(11,f1deltanum,s=s,i=i))}

f1deltad<function(i){sum(kh(x[i][99]-x(i)[1:99],b2[i]))}

f1delta<function(s,i){f1deltan(s,i)/f1deltad(i)}

fd1<-Vectorize(function(y,i){f1delta(y,i)})

curve(fx(x,17),-15,15)
curve(fd1(x,17),-15,15,col=2,add=T)
curve(fd(x,17),-15,15,col=3,add=T)

# for the AR(1)+ARCH(1) use
ARCH<-garchSpec(model = list(ar = 0.6, alpha = 0.7, beta = 0,omega=2))

#rest of the code is the same as previous
AR2<-matrix(NA,nrow=100,ncol=20)
for (i in 1:20)
{
AR2[,i]<-arima.sim(list(order=c(2,0,0),ar=c(s,v)), n=100)
}

varAR2<-function(theta1,theta2){(1-theta2)/((1+theta2)*
(1-theta1-theta2)*(1+theta1-theta2))}

varAR2(s,v)

gamma1<-function(theta1,theta2){(theta1*varAR2(theta1,theta2)/(1-theta2))}
gamma1(s,v)

gamma2<-function(theta1,theta2){theta1*gamma1(theta1,theta2)+
theta2*varAR2(theta1,theta2)}
gamma2(s,v)

sigma22<-function(theta1,theta2){matrix(c(varAR2(theta1,theta2),
gamma1(theta1,theta2),gamma1(theta1,theta2),
varAR2(theta1,theta2)),nrow=2,byrow=T)}
sigma22(s,v)

sigma12<-function(theta1,theta2){matrix(c(gamma1(theta1,theta2),
gamma2(theta1,theta2)),nrow=1,byrow=T)}
sigma12(s,v)

LAST<-function(i){matrix(c(AR2[99,i],AR2[98,i]),nrow=2)}

MEAN<-function(theta1,theta2,i){sqrt(sigma12(theta1,theta2))%^%
solve(sqrt(sigma22(theta1,theta2)))%^%LAST(i)}
MEAN(s,v,i)

SD<-function(theta1,theta2){varAR2(theta1,theta2)-sigma12(theta1,theta2)%^%
solve(sigma22(theta1,theta2))%^%t(sigma12(theta1,theta2))}
SD(s,v)

75
apply(AR2, 2, mean)
apply(AR2, 2, sd)
l1<-c(1:98); n<-100
write.csv(AR2, ”C:/Users/aziz/Desktop/Time_Series/AR2.csv”)
AR2<-as.matrix(read.csv(”C:/Users/aziz/Desktop/Time_Series/AR2.csv”))
AR2<-AR2[, -1]
y<-function(i){AR2[3:100, i]}
z<-function(i){AR2[1:98, i]}
w<-function(i){AR2[2:99, i]}
kh<-function(x, h){(1/h)*exp(-0.5*(x^2/h^2))}
khAR2<-function(l, j, i, h){kh(y(i)[l]-y(i)[j], h)}
nkhAR2<-function(l, j, i, h){(1/h)*exp(-0.5*((w(i)[l]-w(i)[j])^2+(z(i)[l]-z(i)[j])^2)/h^2)}
AR2Newr<-function(l, j, i, h){nkhAR2(l, j, i, h)*khAR2(l, j, i, h)}
AR2Newr1<-function(l, i, h){log((sum(AR2Newr(l, l1, i, h)))-AR2Newr(l, 1, i, h))}
optim(c(1,1), AR2Newr3, control=list(trace=1, maxit=100), i=1)
kb<-function(u, b){(1/b)*exp(-0.5*(u^2/b^2))}
numkb<-function(l, j, i, h){nkhAR2(l, j, i, h)*y(i)[j]}
AR2b0n<-function(l, i, h){numkb(l, l1[l], i, h)}/AR2b0d(l, i, h))
AR2b0d<-function(l, i, h){sum(nkhAR2(l, 1, i, h))}
kk2<-function(l, j, i, h){numkb(l, j, i, h)/(sum(nkhAR2(l, 1, 98, i, h))-1)}
mhat2<-function(l, i, h){sum(kk2(l, 1, 98, i, h))-y(1)[l]*nkhAR2(l, 1, i, h)}
mhat2<-function(l, i, h){sum(AR2b0n(l, i, h)/AR2b0d(l, i, h))}
ehat2<-function(l, i, h){y(i)[l]-mhat2(l, i, h)}
ehat2a<-function(l, i, h){(ehat2(l, i, h))^2}
sse<-function(i, h){(1/n)*Reduce(”+”, lapply(1:98, ehat2a, i=i, h=h))}
optim(1,sse,method="Brent",lower=0.05,upper=50,i=7)

AR2Newr2<-function(i,h){(1/n)*Reduce(”+”,lapply(ll,AR2Newr1,i=i,h=h))}

AR2Newr3<-function(i,h){-AR2Newr2(i,h)}

AR2values<-read.csv("C:/Users/aziz/Desktop/Time_Series/AR2values.csv")
h1AR2<-AR2values[,1]
h2AR2<-AR2values[,2]
b0AR2<-AR2values[,4]

ww<-function(l,i){ehat2(l,i,b0AR2[i])}

kbAR2<-function(l,j,i,h){kb(ww(l,i)-ww(j,i),h)}

AR2Bewr<-function(l,j,i,h){nkhAR2(l,j,i,h[2])*kbAR2(l,j,i,h[1])*(1/(n-1))}

AR2Bewr1<-function(l,i,h){log(Reduce("+",lapply(ll,AR2Bewr,l=l,i=i,h=h))
-AR2Bewr(l,l,i,h))}

AR2Bewr2<-function(l,i,h){(1/n)*Reduce("+",lapply(1:98,AR2Bewr1,i=i,h=h))}

AR2Bewr3<-function(i,h){-AR2Bewr2(i,h)}

optim(c(1,1),AR2Bewr3,control=list(trace=1,maxit=100),i=7)

mconst<-1/sqrt(2*pi)

fx<-function(x,i){dnorm(x,MEAN(s,v,i),SD(s,v))}

fx1<-function(x,i){dnorm(x,MEAN(s,v,i),sqrt(varAR2(s,v)))}

AR2values<-read.csv("C:/Users/aziz/Desktop/Time_Series/AR2values.csv")

AR2h1<-AR2values[,1]

AR2h2<-AR2values[,2]

AR2b1<-AR2values[,5]

AR2b2<-AR2values[,6]

fdeltan<-function(u,i){mconst*(sum(kh(u-y(i),AR2h1[i])*sqrt((z(i)[98]-z(i))^2+(w(i)[98]-w(i))^2),AR2h2[i])/))}

fdeltad<-function(i){sum(kh(sqrt((z(i)[98]-z(i))^2+77}}
(w(i)[98]−w(i)\^2,AR2h2[i]))

fdelta<-function(u,i){fdeltan(u,i)/fdeltad(i)}
fd<-Vectorize(function(u,i){fdelta(u,i)})
curve(fx(x,9),-10,10)
curve(fd(x,9),-10,10,add=T,col=2)
curve(fx1(x,9),-10,10,add=T,col=4)
mhat2n<-function(u,i,h){sum(kh(sqrt((u[1]−z(i))^2+(u[2]−w(i))^2),h)*y(i))}
mhat2d<-function(u,i,h){sum(kh(sqrt((u[1]−z(i))^2+(u[2]−w(i))^2),h))}
mhat2x<-function(u,i,h){mhat2n(u,i,h)/mhat2d(u,i,h)}
mm2hat<-function(i){mhat2x(c(z(i)[98],w(i)[98]),i,b0AR2[i])}
ehat2<-function(1,i,h){y(i)[1]−mhat2(1,i,h)}
e2<-function(1,i){ehat2(1,i,b0AR2[i])}
f1deltanum<-function(s,i,l){mconst*(sum(kh(sqrt((s−mm2hat(i)−e2(1,i),AR2b1[i])
*sqrt((z(i)[98]−z(i)[1])^2+(w(i)[98]−w(i)[1])^2),AR2b2[i])))}
f1deltan<-function(s,i){Reduce("+",lapply(ll,f1deltanum,s=s,i=i))}
f1deltad<-function(i){sum(kh(sqrt((z(i)[98]−z(i))^2+(w(i)[98]−w(i)[1])^2),AR2b2[i]))}
f1delta<-function(s,i){f1deltan(s,i)/f1deltad(i)}
fd1<-Vectorize(function(y,i){f1delta(y,i)})
curve(fx(x,4),-10,10)
curve(fd(x,4),-10,10,add=T,col=2)
curve(fd1(x,4),-10,10,add=T,col=3)

#where (s,v)=(0.3,0.5),(0.2,0.15) and (0.5,0.45)

x<-read.csv("C:/Users/aziz/Desktop/Time_Series/NYSE.csv")
y<-read.csv("C:/Users/aziz/Desktop/Time_Series/TSX.csv")
\[
x_1 \leftarrow x[1:750,2] \\
x_2 \leftarrow y[1:750,2] \\
y_1 \leftarrow x[2:751,2] \\
y_2 \leftarrow y[2:751,2] \\
l \leftarrow c(1:750); n \leftarrow 752 \\
kh \leftarrow \text{function}(x, h) \{ (1/h) \exp(-0.5 \times (x^2/h^2)) \} \\
NYSENewr \leftarrow \text{function}(1, j, h) \{ kh(x[1:1] - x[1:j], h[2]) \times kh(y[1:1] - y[1:j], h[2]) \times 1/(n-1) \} \\
NYSENewr1 \leftarrow \text{function}(1, h) \{ \log(\text{sum}(\text{NYSENewr}(1, ll, h)) - \text{NYSENewr}(1, 1, h)) \} \\
NYSENewr2 \leftarrow \text{function}(h) \{ (1/n) \times \text{Reduce}("+", \text{lapply}(ll, \text{NYSENewr1}, h=h)) \} \\
NYSENewr3 \leftarrow \text{function}(h) \{ -\text{NYSENewr2}(h) \times \text{kh}(363, ll, \text{comboh2}) \} \\
\text{optim}(c(50,50), \text{NYSENewr3}, \text{control}=\text{list}(\text{trace}=1, \text{maxit}=300)) \\
NYSeh[1] \leftarrow 55.24414 \\
NYSeh[2] \leftarrow 45.70801 \\
NYSeh[\text{value}] \leftarrow 12.22893 \\
\]

\[
\text{kbnyse} \leftarrow \text{function}(1, j, b) \{ kb(x[1:1] - x[1:j], b) \times y[1:j] \} \\
\text{kbnyse} \leftarrow \text{function}(1, b) \{ \text{kbnyse}(1, ll, b) \} \\
\text{kbnyse} \leftarrow \text{function}(1, j, b) \{ kb(x[1:1] - x[1:j], b) \} \\
\text{mhatnyse} \leftarrow \text{function}(1, b) \{ (\text{sum}(\text{kbnyse}(1, ll, b)) - y[1:1]/b) \\
/ (\text{sum}(\text{kbnyse}(1, ll, b)) - 1/b) \} \\
\text{ehatnyse} \leftarrow \text{function}(1, b) \{ y[1:1] - \text{mhatnyse}(1, b) \} \\
\text{ehatnysea} \leftarrow \text{function}(1, b) \{ (\text{ehatnyse}(1, b))^2 \} \\
\text{ssenyse} \leftarrow \text{function}(b) \{ (1/n) \times \text{Reduce}("+", \text{lapply}(ll, \text{ehatnysea}, b=b)) \} \\
\text{optim}(1, \text{ssenyse}, \text{method}="Brent", \text{lower}=1, \text{upper}=500) \\
b0nyse \leftarrow 49.27849 \\
\text{wwtsx} \leftarrow \text{function}(1) \{ \text{ehttsx}(1, b0tsx) \} \
\]

79
wwnyse <- function(l) { ehatnyse(l, b0nyse) }

kn <- function(l, j, b) { kb(wwnyse(l) - wwnyse(j), b) }

Bewrnyse <- function(l, j, b) { kb(x1[l] - x1[j], b[2])*kn(l, j, b[1])*(1/(n-1)) }

Bewrnyse1 <- function(l, j, b) { log(Reduce(“+”, lapply(l1, Bewrnyse, l=l, b=b) - Bewrnyse(l, l, b)) ) }

Bewrnyse2 <- function(b) { (1/n)*Reduce(“+”, lapply(l1, Bewrnyse1, b=b) ) }

Bewrnyse3 <- function(b) { -Bewrnyse2(b) }

optim(c(50,50), Bewrnyse3, control=list(trace=1, maxit=100))

NYSEb1 <- 39.93191
NYSEb2 <- 140.16763
NYSEbvalue <- 12.19342

kkx <- function(l, j, h) { (1/h^2)*exp(-0.5*((x1[l] - x1[j])^2 + (x2[l] - x2[j])^2)/h^2) }

Newrnyse <- function(l, j, h) { kkx(l, j, h[2])*kh(y1[l] - y1[j], h[1])*(1/(n-1)) }

Newrnyse1 <- function(l, h) { log((sum(Newrnyse(l, ll, h))) - Newrnyse(l, l, h)) }

Newrnyse2 <- function(h) { (1/n)*Reduce(“+”, lapply(l1, Newrnyse1, h=h)) }

Newrnyse3 <- function(h) { -Newrnyse2(h) }

optim(c(50,50), Newrnyse3, control=list(trace=1, maxit=300))

h1NYSE <- 59.90662
h2NYSE <- 58.93257
hvalueNYSE <- 18.07153

khll <- function(l, j, h) { (1/h^2)*exp(-0.5*((y1[l] - y1[j])^2 + (y2[l] - y2[j])^2)/h^2) }

aa <- (1/sqrt(2*pi))^2

comboNewr <- function(l, j, h) { kkx(l, j, h[2])*khll(l, j, h[1])*(1/(n-1)) }

comboNewr1 <- function(l, h) { log((sum(comboNewr(l, ll, h))) - comboNewr(l, l, h)) }

comboNewr2 <- function(h) { (1/n)*Reduce(“+”, lapply(l1, comboNewr1, h=h)) }

comboNewr3 <- function(h) { -comboNewr2(h) }

80
\begin{verbatim}
optim(c(50,50),comboNewr3, control=list(trace=1,maxit=300))
h1combo<-63.48577
h2combo<-62.93158
hvaluecombo<-23.13319

khl2<-function(a,b,j,h){(1/h^2)*exp(-0.5*((a-y1[j])^2+(b-y2[j])^2)/h^2)}
fnc<-function(a,b){aa*(sum(kx(length(ll),ll,h2combo)*khl2(a,b,ll,h1combo)))/sum(kx(length(ll),ll,h2combo))}
f<-Vectorize(function(a,b){fnc(a,b)})
margnyse<-function(a){integrate(fc,min(x2),max(x2),a=a)}$value
mnyse<-Vectorize(function(a){margnyse(a)})
fnnyse<-function(y){sum(kx(x1[length(ll)]-x1[ll],NYSEh2)*kx(y-y1[ll],NYSEh1))/sum(kx(x1[length(ll)]-x1[ll],NYSEh2))}
fnyse<-Vectorize(function(y){fnnyse(y)/sqrt(2*pi)})

curve(fnyse,10500,11200)
nymhatn<-function(u,h){sum(kx(u-x1[ll],h)*y1[ll])}
nymhatd<-function(u,h){sum(kx(u-x1[ll],h))}
nymhatx<-function(u,h){nymhatn(u,h)/nymhatd(u,h)}
nymhat<-nymhatx(x1[length(ll)],b0nyse)
ehatny<-function(1,b){y1[1]-mhatnyse(1,b)}
nysehat<-function(1){ehatny(1,b0nyse)}
flnysenum<-function(s,l){(1/sqrt(2*pi))*(sum(kx(s-nymhat-nysehat(l),NYSEb1)*kx(x1[length(ll)]-x1[1],NYSEb2)))}
flnysen<-function(s){Reduce("+",lapply(11,flnysenum,s=s))}
flnysed<-sum(kx(x1[length(ll)]-x1[11],NYSEb2))
flnyse<-function(s){flnysen(s)/flnysed}
\end{verbatim}
fnys2 <- Vectorize(function(s){f1nyse(s)})
curve(fnys2,10500,11200,col=3)
curve(fnys,10500,11200,add=T,col=2)
fnysn <- function(y){sum(kkx(length(l),l,h2NYSE) * kh(y - y1[l],h1NYSE))/sum(kkx(length(l),l,h2NYSE))}
curve(fnys,10500,11200,col=3)
curve(fnys,10500,11200,add=T,col=2)
curve(fnysn,10500,11200,add=T,col=4)
curve(mnys,10500,11200,add=T)
#same is done for TSX

DATA <- read.csv("C:/Users/aziz/Desktop/Time-Series/Weather.csv")
DATA <- as.numeric(DATA)
DATA <- DATA[1:365,]
Windspd <- DATA[,2]
Prec <- DATA[,3]
MinTemp <- DATA[,4]
MaxTemp <- DATA[,5]
Pressure <- DATA[,6]
Temp <- DATA[,7]
Data <- DATA[,-1]
pairs(Data)
panel.cor <- function(x, y, digits=2, prefix="", cex.cor) {
    usr <- par("usr"); on.exit(par(usr))
par(usr = c(0, 1, 0, 1))
r <- abs(cor(x, y, use="complete.obs"))
txt <- format(c(r, 0.123456789), digits=digits)[1]
txt <- paste(prefix, txt, sep="")
if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)*r
text(0.5, 0.5, txt, cex = cex.cor)
}

panel.hist <- function(x, ...) {
usr <- par("usr"); on.exit(par(usr))
par(usr = c(usr[1:2], 0, 1.5))
h <- hist(x, plot = FALSE)
breaks <- h$breaks; nB <- length(breaks)
y <- h$counts; y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)
}
pairs(Data, upper.panel=panel.cor, diag.panel=panel.hist)
x2<Temp[1:363]
y2<Temp[2:364]
x1<Windspd[1:363]
y1<Windspd[2:364]
x3<Pressure[1:363]
y3<Pressure[2:364]
l<-c(1:363); n<365
kh<function(x, h){(1/h)*exp(-0.5*(x^2/h^2))}
onetemp<function(1, j, h){kh(x2[j]-x2[j], h[2])*kh(y2[j]-y2[j], h[1])}
onetemp1<function(1, h){log(sum(onetemp(1, l1, h))-1/(h[1]*h[2]))}
onetemp2 <- function(h) {
  (1/n) * Reduce("+", lapply(ll, onetemp1, h=h))
}

onetemp3 <- function(h) {-onetemp2(h)}

optim(c(50, 50), onetemp3, control=list(trace=1, maxit=300))

temph1 <- 1.741543

temph2 <- -1.657396

temphvalue <- -1.344064

kb <- function(x, b) {1/b * exp(-0.5*(x^2/b^2))}

mtt <- function(l, j, b) {kb(x2[l] - x2[j], b) * y2[j]}

mtd <- function(l, j, b) {kb(x2[l] - x2[j], b)}

mnt <- function(1, b) {mtt(1, ll[-1], b)}

mtd2 <- function(1, b) {sum(mtd(1, ll[-1], b))}

kktemp <- function(1, b) {mnt(1, b) / mtd2(1, b)}

mhattemp <- function(1, b) {sum(kktemp(1, b))}

ehattemp <- function(1, b) {y2[1] - mhattemp(1, b)}

ehattemp2 <- function(1, b) {(ehattemp(1, b))^2}

ssetemp <- function(b) {(1/n) * Reduce("+", lapply(ll, onetemp2, b=b))}

optim(0.0001, ssetemp, method="Brent", lower=0.00001, upper=50)

b0temp <- 2.114106

wwtemp <- function(l) {ehattemp(l, b0temp)}

kw <- function(l, j, b) {kb(wwtemp(1) - wwtemp(j), b)}

Bewrtemp <- function(l, j, b) {kb(x2[l] - x2[j], b[2]) * kw(1, j, b[1])}

Bewrtemp1 <- function(l, b) {log(sum(Bewrtemp(1, ll[-1], b)))}

Bewrtemp2 <- function(b) {(1/n) * Reduce("+", lapply(ll, Bewrtemp1, b=b))}

Bewrtemp3 <- function(b) {-Bewrtemp2(b)}

optim(c(1, 1), Bewrtemp3, control=list(trace=1, maxit=300))

tempb1 <- -5.215444
tempb2<-2.269433

tempbvalue<- -0.7038377

khx<-function(1,j,h){(1/h^3)*exp(-0.5*((x1[l]-x1[j])^2+(x2[l]-x2[j])^2+(x3[l]-x3[j])^2)/h^2)}
khy<-function(1,j,h){(1/h^3)*exp(-0.5*((y1[l]-y1[j])^2+(y2[l]-y2[j])^2+(y3[l]-y3[j])^2)/h^2)}

AR1Newr<-function(1,j,h){khx(1,j,h[2])*khy(1,j,h[1])*(1/(n-1))}

AR1Newr1<-function(1,h){log(((sum(AR1Newr(1,l1,h)))-AR1Newr(1,1,h)))}

AR1Newr2<-function(h){(1/n)*Reduce(“+”,lapply(l1,AR1Newr1,h=h))}

AR2Newr2<-function(h){(1/n)*Reduce(“+”,lapply(l1,AR2Newr1,h=h))}

AR1Newr3<-function(h){-AR1Newr2(h)}

optim(c(50,50),AR1Newr3,control=list(trace=1,maxit=300))

comboh1<-1.247444

comboh2<-1.655643

combohvalue<-5.711223

fntemp<-function(y){(sum(kh(x2[363]-x2[1:363],temph2)*kh(y-y2[1:363],temph1)))/sum(kh(x2[363]-x2[1:363],temph2))}

fnntemp<-function(y){fntemp(y)/sqrt(2*pi)}

ftemp<-Vectorize(function(y){fnntemp(y)})

optim(100,AR1Newr3,control=list(trace=1,maxit=300))

curve(ftemp(x),y2[363]-15,y2[363]+15)

tempmhatn<-function(u,h){sum(kh(u-x2[11],h)*y2[11])}

tempmhatd<-function(u,h){sum(kh(u-x2[11],h))}

tempmhatx<-function(u,h){tempmhatn(u,h)/tempmhatd(u,h)}

tempmhat<-tempmhatx(x2[363],b0temp)

ehattemp<-function(1,b){y2[1]-mhattemp(1,b)}
temp<scalefunction(1)\{ehattemp(1,b0temp)\}

fltempnum<scalefunction(s,1)\{(1/sqrt(2*pi))*\{sum(kh(s-tempmhat-tempehat(1),tempb1)*kh(x2[363]-x2[l],tempb2))\}\}

fltempn<scalefunction(s)\{Reduce("+",lapply(ll,fltempnum,s=s))\}

fltempd<sum(kh(x2[363]-x2[l],tempb2))

fltemp<scalefunction(s)\{fltempn(s)/fltempd\}

ftemp2<Vectorize(scalefunction(s)\{f1temp(s)\})

curve(f1temp(x),y2[363]-15,y2[363]+15,col=2)
curve(ftemp2(x),y2[363]-15,y2[363]+15,col=3,add=T)

# the same is done for wind and pressure variables