

Spring 5-2018

Nonparametric Estimation of Time Series Volatility Model Estimation

Teng Tu

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds



Part of the [Statistical Models Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Tu, Teng, "Nonparametric Estimation of Time Series Volatility Model Estimation" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1496.

https://openscholarship.wustl.edu/art_sci_etds/1496

This Thesis is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Mathematics

Nonparametric Estimation of Time Series Volatility Model Estimation

by

Teng Tu

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2018

St. Louis, Missouri

© 2018, Teng Tu

Table of Contents

	Page
List of Figures	iii
List of Tables	iv
Acknowledgments	v
Abstract of the Thesis	vi
1 Introduction	1
2 Estimation based on two-lag difference	4
2.1 Cross-validation	4
2.2 Local Linear Regression	5
2.3 Selection of the bandwidth	7
2.4 Estimation approach based on the two-lag difference	8
2.5 Simulation result	12
3 Estimation based on general quadratic forms	15
3.1 Basic idea and algorithm	15
3.2 Simulation result	19
4 Conclusion	24
References	25

List of Figures

Figure	Page
3.1 Estimation of variance where $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$	22
3.2 Estimation of variance where $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$	23

List of Tables

Table		Page
2.1	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$	14
2.2	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$	14
3.1	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$ and the parameter (i) $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$	20
3.2	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ and the parameter (i) $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$	20
3.3	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$ and the parameter (ii) $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$	21
3.4	Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ and the parameter (ii) $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$	21

Acknowledgments

First and foremost, I would like to sincerely thank Prof. José E Figueroa-López for his suggestions and ideas when I working on this thesis. I cannot imagine without his guidance and support I can accomplish this thesis.

In addition, for all the faculty, staff and friends in our Math Department, thanks for your kindness and encouragement. You make my life a wonderful experience in our Math Department.

Last but not least, I would like to thank my parents for their love and support, you give me chance to study in Math Department and have an amazing experience.

Teng Tu

Washington University in St. Louis

May 2018

Abstract of the Thesis

Nonparametric Estimation of Time Series Volatility Model Estimation

by

Master of Arts in Statistics,

Washington University in St. Louis, 2018.

Professor José E Figueroa-López, Chair

In this article we consider two estimation methods of a non-parametric volatility model with autoregressive error of order two. The first estimation method based on the two-lag difference [1]. To get a better result, we consider the second approach based on the general quadratic forms [2]. For illustration, we provided several data sets from different simulation models to support the procedures of both two methods, and prove that the second approach can make a better estimation.

Keywords and phrases: autoregressive error process, two-lag difference, general quadratic forms.

1. Introduction

This paper considers the estimation of a time series process with a time-dependent conditional variance function and serially dependent errors.

Considering T observations $\{(x_t, y_t)\}_{t \in \{1, \dots, T\}}$ generated by the following model:

$$y_t = \sigma_t v_t \quad \sigma_t = \sigma(x_t) \quad (1.1)$$

$$v_t = \sum_{j=1}^p \phi_j v_{t-j} + \epsilon_t \quad (1.2)$$

for $t = 2, \dots, T$, where $\{\epsilon_t : -\infty < t < \infty\}$ are independent identically distributed random variable with mean 0 and variance 1. The autoregressive order is a fixed known integer $p > 0$, in this paper, we focus more on $p=2$ model. We also assume that x_t 's form an increasing equally spaced sequence on the interval $[0, 1]$. The model (1.1)-(1.2) can be viewed as a nonparametric regression model with the mean function identically equal to zero and scaled autoregressive time series errors v_t . It can be written as

$$y_t = \phi_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1} + \phi_2 \sigma_t \sigma_{t-2}^{-1} y_{t-2} + \sigma_t \epsilon_t \quad (1.3)$$

For the case that $p=1$, Dahl and Levine (2006) [3] studied a method based on the two-lag difference statistics:

$$\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}} \quad (1.4)$$

to do the estimation. The detail of it is explained in the Section 4 Chapter 2.

Nonparametric regression model with time series errors has a long history. Lin et al. (1999) [4] considered the estimation of a regression models of the form

$$y_t = g(x_t) + z_t, \quad (1.5)$$

where, (x_t, y_t) 's ($1 \leq t \leq n$) are observed data, $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is a unknown smooth regression function, and the model has a constrain that $\{x_t\}$ is a covariate process independent of the stationary error $\{z_t\}$.

Hall and Keilegom (2003) [5] estimated a general nonparametric model by using difference-based method for inference in nonparametric regression with independent errors. The model they considered has the form

$$y_t = \sigma(x_t) + v_t, \quad (1.6)$$

where, (x_t, y_t) 's ($1 \leq t \leq n$) are observed data, σ is a smooth function and the error process v_t is the same as the error process v_t mentioned in (1.2). Comparing the model mentioned in (1.3), this model did not constrain that $\{x_t\}$ should be a covariate process, instead, they consider that $\{x_t\}$ is an increasing sequence on the interval $[0, 1]$. They used the difference operator defined as $(D_j y)_t = y_t - y_{t-j}$ to estimate the model (1.5) and then estimate the covariance structure $\gamma(j) = cov(v_t, v_{t-j})$ by the Yule-Walker equations to link autoregressive structure to covariance.

Shao and Yang (2011) [6] considered the Yule-Walker estimator of the same model mentioned in (1.5) by B-splines method.

Considering that the model (1.1) discussed in this paper rescales the AR(2) error process by the conditional variance of the unobserved process $\sigma(x_t)$, it has a more general correlation structure than the model (1.5).

As mentioned above, this paper includes an account of the work of Dahl and Levine (2006) [3] who treated the case $p = 1$ by the two-lag difference statistics. The model considered in this paper which based on the Figueroa-López and Levine (2013) [1] and Figueroa-López(2013b) [2] extend the method for considering the case for $p = 2$ and show that the method can also work in this situation.

The rest of the paper is organized as follows. In Chapter 2, we present our estimation method based on two-lag differences for AR(2) case and the simulation results are also shown. The estimation approach based on general quadratic forms and its simulations are also presented in Chapter 3.

2. Estimation based on two-lag difference

The first section talks about the method of cross-validation which is used to decrease the problem of overfitting in estimation, and the second section explains the basic idea of Local Linear Regression which is a popular method to estimate non-parametric objects. The third section talks about the idea of selecting the bandwidth in Local Linear Regression based on the cross-validation method. And the next section explain the estimation of model (1.1)-(1.2) based on two-lag difference which is mentioned in Figueroa-López and Levine (2013b) [1]. In the final section, we present our simulations based on this estimation method, and show that their performance are good.

2.1 Cross-validation

Before talking about statistical models, we first talk about a method that help us to do the model selection. Cross-validation is a validation technique for model evaluation. In the model evaluation, we always want to choose a statistical learning method which has a low test error. Unfortunately, it has a problem that we could not know what the prediction results are for the new data we have not seen, so it is really hard for us to calculate the test error. Cross-validation is one way to overcome this problem by separating the data into several parts instead of using the entire dataset to do the evaluation. In the training process, it removed some parts of the data to train the model, when the training is done, the testing process will use the remaining parts of the data

to evaluate the performance of the model. There are several kinds of Cross-validation method, in this article, we used the one which called the K-fold cross validation method. Take this case as an example: if we have some data $(x_1, y_1), \dots, (x_n, y_n)$, and want to use this data to estimate a model using an estimator that depends on an unknown parameter α , i.e. $\hat{\beta}(\alpha)$. The K-fold cross-validation works as the following way: it separated this dataset into K parts which almost have the same size. We used $K - 1$ parts of the data to train the model with the parameter α . Let $\hat{\beta}^{-k}(x, \alpha)$ be the fitted function, and then compute the error in predicting the k th part, which we call it testing part:

$$E_k(\alpha) = \sum_{i=1}^{n_k} (y_i - \hat{\beta}(x_i, \alpha))^2, \quad (2.1)$$

where x_i is the i th observation in the k th part of data, y_i is the response of the i th observation in the k th part of data, n_k is the number of observations in the k th part of data. Then we repeat this step for K times and guarantee that all of these parts can be the testing part, and then get the cross-validation error:

$$CV_K(\alpha) = \frac{1}{K} \sum_{k=1}^K E_k(\alpha). \quad (2.2)$$

We do this for different values of α and the one which minimize the $CV(\alpha)$ is the value we choose to estimate the parameters of the model. For further reference see [7].

2.2 Local Linear Regression

Local Linear Regression is a common method to estimate a non-parametric model.

Consider a general non-parametric model

$$y = f(x) + \epsilon \quad (2.3)$$

in which $x = (x^1, \dots, x^p) \in \mathbb{R}^p$, $y \in \mathbb{R}$, $f(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}$ is a smooth and unknown structure, $\epsilon \in \mathbb{R}$ is a independent identically distributed with mean 0 and variance 1.

Since the function $f(\cdot)$ is smooth enough, it will look linear in small regions of input space (which indicate all the possible inputs). Suppose that we consider points in input space nearby x_0 . The model could be approximated as:

$$y = \beta_0[x_0] + \sum_{j=1}^p \beta_j[x_0] \cdot (x^j - x_0^j) + \epsilon \quad (2.4)$$

where x^j is the j -th dimension of $x = (x^1, \dots, x^p)$,

for x near x_0 should satisfy $\|x - x_0\| \leq h$, i.e. the distance between x and x_0 should be equal of less than h , the h is called the bandwidth of the model, $[x_0]$ are used to represent the fact that the value of β will vary for different values of x_0 .

The local linear regression to estimate $\hat{\beta}_0[x_0], \hat{\beta}[x_0]$ is given by minimizing the following function:

$$\operatorname{argmin}_{\beta_0, \beta} \sum_{i=1}^n K(\|x_i - x_0\|/h) \cdot (y_i - \beta_0 - (x_i - x_0)' \beta)^2 \quad (2.5)$$

In R, we can use the function [8]:

```
loess(formula, span, deg, loess.control(surface, trace.hat))
```

to do the local lineal regression.

For the parameters shown in the R function above, we have:

- *formula* is a formula specifying the numeric response and one to four numeric predictors,
- *span* is the bandwidth,
- *deg* is the degree of the polynomials to be used,

- *surface* is a parameter indicates should the fitted surface be computed exactly ("direct") or via interpolation from a kd tree ("interpolate"),
- *trace.hat* means that compute the trace of the smoother matrix exactly("exact") or approximately("approximate"). It is recommended to use the approximation for more than about 1000 data points.

2.3 Selection of the bandwidth

As mentioned above, the selection of the bandwidth h in the local linear regression method is a key part to perform the regression. If we choose a large h , then the local linear assumption is not accurate. On the other hand, if we choose a very small h , the estimation will not be accurate because only a few data points will be considered.

In our estimation, we select the proper bandwidth h based on the 10-fold cross validation.

Firstly, we vary the bandwidth h from 0.05 to 1 by 0.05. For a particular bandwidth h_j we separate the data into ten parts, then use nine parts of the data with bandwidth h_j to train the model $\hat{\beta}^{-k}(x, h_j)$, and use the remaining one part to compute the evaluate error (2.1). Then we repeat this for ten times and guarantee that all of these ten parts can be the testing part, and then use (2.1) to get the cross-validation error $CV(h_j)$ for bandwidth h_j .

The last step is that we compare all the cross-validation error based on different bandwidth h , the one which gives the smallest cross-validation error is the bandwidth we would like to choose in our estimation.

2.4 Estimation approach based on the two-lag difference

As mentioned above, Hall and Keilegom (2003) [3] treated the case $p = 1$ by the two-lag difference statistics. and Figueroa-López and Levine (2013b) [1] showed that this method also works for the case $p = 2$. Now, we exhibit more details about that paper. Firstly, we consider the model (1.1)-(1.2) based on the AR(1) model, i.e.

$$y_t = \sigma_t v_t, \sigma_t = \sigma(x_t),$$

$$v_t = \phi_1 v_{t-1} + \epsilon$$

It is known that for a AR(1) model:

$$v_t = \phi_1 v_{t-1} + \epsilon,$$

the covariance $\gamma_1 = cov(v_t, v_{t-1})$ and $\gamma_2 = cov(v_t, v_{t-2})$ have the relationship: $\gamma_2 = \phi_1 \gamma_1$.

Combining this with the two-lag statistics $\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}}$, we could get

$$E\eta_t^2 = E\frac{\sigma^2}{2}(Ev_t^2 + Ev_{t-2}^2 - 2Ev_t v_{t-2}) = \sigma^2(\gamma_0 - \gamma_2) = \sigma^2, \quad (2.6)$$

where $\sigma^2 = \gamma_0 = var(v_t)$.

According to this, we can think that η_t^2 can be used to develop a consistent estimator for a non-constant function σ_t^2 . For a general σ_t and under $x_t = \frac{t}{T}, t = 0, \dots, T$, we will have

$$E\eta_t^2 = \frac{1}{2}(\sigma_t^2 \gamma_0 + \sigma_{t-2}^2 \gamma_0 - 2\sigma_t \sigma_{t-2} \gamma_2)$$

When T is very large and the function σ is smooth we would have $\sigma_{t-2} \approx \sigma_t$, and it is naturally to find that this expression can be accurately approximated by σ_t^2 . That is, we will have:

$$E\eta_t^2 \approx \sigma_t^2 = \sigma^2(x_t).$$

Because of this, we can change the original problem (1.1) into a non-parametric regression problem with the form:

$$\eta_t^2 = \sigma^2(x_t) + \tilde{\epsilon}_t, \quad (2.7)$$

where $\tilde{\epsilon}_t$ are approximately centered random errors. Figueroa-López and Levine (2013b) [3] proposes a weighted least square estimator (WLSE). Concretely, noting that (1.3) with $\phi_2 = 0$ implies

$$\sigma_t^{-1}y_t = \phi_1\sigma_{t-1}^{-1}y_{t-1} + \epsilon_t, \quad t = 2, \dots, T, \quad (2.8)$$

a natural estimator for ϕ_1 is given by:

$$\begin{aligned} \hat{\phi}_1 &:= \operatorname{argmin}_{\phi_1} \frac{1}{T} \sum_{t=2}^T (\hat{\sigma}_t^{-1}y_t - \phi_1\hat{\sigma}_{t-1}^{-1}y_{t-1})^2 \\ &= \left(\frac{1}{T} \sum_{t=2}^T \hat{\sigma}_{t-1}^{-2}y_{t-1}^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=2}^T \hat{\sigma}_t^{-1}\hat{\sigma}_{t-1}^{-1}y_t y_{t-1} \right) \end{aligned} \quad (2.9)$$

However, the method described above does not work for the case $p > 1$.

To simplify this problem, we consider the case where $p = 2$, i.e.:

$$y_t = \sigma_t v_t \quad \sigma_t = \sigma(x_t) \quad (2.10)$$

$$v_t = \phi_1 v_{t-1} + \phi_2 v_{t-2} + \epsilon_t \quad (2.11)$$

Therefore, we want to find that whether there exists other linear statistics

$$\eta_t := \sum_{i=1}^m a_i y_{t-i} \quad (2.12)$$

such that $E\eta_t^2 \approx \sigma_t^2$.

Figueroa-López and Levine (2013b) [1] shows the following proposition:

Proposition 2.1 Suppose that $\phi_2 = 0$ and $\sigma(\cdot) \equiv \sigma \in \mathbb{R}_+$ for a unknown positive constant. Then, if

$$E\eta_n^2 = \sigma^2,$$

for any $\phi_1 \in (-1, 1)$, there exists a $0 \leq k \leq m - 2$ such that

$$a_k = \pm \frac{1}{\sqrt{2}}, a_{k+2} = \mp \frac{1}{\sqrt{2}}, \forall i \neq k, k + 2.$$

This result shows that when we want to get a $E\eta_t^2$ which is independent of ϕ_1 based on the linear statistics (2.12) with $a_0 \neq 0$, the only one statistics could satisfy this requirement is the two-lag difference statistics:

$$\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}}.$$

For a general AR(2) innovation process and for $\sigma(\cdot) \equiv \sigma$, the equation (2.6) can be simplifies as:

$$E\eta_t^2 = \sigma^2(\gamma_0 - \gamma_2) = \frac{\sigma^2}{1 + \phi_2}.$$

What's more, it can be deduced that

$$\begin{aligned} \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 = \frac{\phi_1^2 + (1 - \phi_2)\phi_2}{1 - \phi_2}\gamma_0, \\ \gamma_0 - \gamma_2 &= \gamma_0\left(1 - \frac{\phi_1^2 + (1 - \phi_2)\phi_2}{1 - \phi_2}\right) = \frac{1}{1 + \phi_2}\gamma_0 \end{aligned}$$

Then, as the deduction in the case $p = 1$, when the function $\sigma(\cdot)$ is general enough and T is large enough, under a fixed design $x_t = \frac{t}{T}, t = 0, \dots, T$, we expect that the following equation will hold

$$E\eta_t^2 \approx \frac{\sigma_t^2}{1 + \phi_2},$$

and similarly, we expect to estimate σ_t^2 up to a constant. Because of this, as in the $p = 1$ case, WLSE will also suffice to estimate the parameter ϕ_1 and ϕ_2 .

Because of this, if we suppose that we know the variance function σ_t and let $\bar{y}_t := \sigma_t^{-1}y_t$,

according to the relationship (1.3), we could estimate (ϕ_1, ϕ_2) by the WLSE:

i.e.

$$\bar{\phi}_2 := (\bar{A}^2 - \bar{B}^2)^{-1}(\bar{A}\bar{C} - \bar{B}^2), \bar{\phi}_1 = \bar{A}^{-1}\bar{B}(1 - \bar{\phi}_2). \quad (2.13)$$

where, $\bar{A} = \sum_{t=4}^T \bar{y}_t^2$, $\bar{B} := \sum_{t=4}^T \bar{y}_t \bar{y}_{t-1}$, $\bar{C} := \sum_{t=4}^T \bar{y}_t \bar{y}_{t-2}$. It seemly that this estimators will not work when σ_t is unknown. However, if we can notice that if we used $c\sigma_t$ for any constant c which is independent of t to replace σ_t in \bar{y}_t , these estimators will not change. Base on this fact, we can have the following algorithm:

1. Estimate the function $\nu^2(x) := \frac{\sigma^2(x)}{1+\phi_2}$ using a non-parametric method for the non-parametric regression

$$\eta_t^2 = \nu^2(x_t) + \tilde{\epsilon}_t, t = 4, \dots, T$$

Let $\hat{\nu}_t = \hat{\nu}(x_t)$ be the resulting estimator.

2. Standardize the observations $\hat{y}_t := \hat{\nu}_t^{-1}y_t$ and then estimate (ϕ_1, ϕ_2) via the WLSE:

$$\hat{\phi}_2 := (A^2 - B^2)^{-1}(AC - B^2), \hat{\phi}_1 = A^{-1}B(1 - \hat{\phi}_2) \quad (2.14)$$

with $A = \sum_{t=4}^T \hat{y}_t^2$, $B := \sum_{t=4}^T \hat{y}_t \hat{y}_{t-1}$, $C := \sum_{t=4}^T \hat{y}_t \hat{y}_{t-2}$.

3. Estimate $\hat{\sigma}_t^2 := \sigma^2(x_t)$ by

$$\hat{\sigma}_t^2 := (1 + \hat{\phi}_2)\hat{\nu}_t^2 \quad (2.15)$$

2.5 Simulation result

In this part, we run a simulation based on the model (1.1) and (1.2) and solve the model by the algorithm mentioned above. In order to do this, we consider the following variance function

$$(a) \quad \sigma_t^2 = (x_t + 0.1)^2, \quad (b) \quad \sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2.$$

with $x_t = \frac{t}{T}$ for $t = 0, \dots, T$.

We can use the following equation to compute the MSE to check the performance of the estimator (2.13):

$$MSE(\hat{\sigma}) := \frac{1}{M} \sum_{i=1}^M \frac{1}{T} \sum_{t=1}^T (\hat{\sigma}_{t,i}^2 - \sigma_t^2)^2,$$

where $\hat{\sigma}_{t,i}$ is the estimated variance function in the i^{th} simulation and M is the number of simulations. And we use local linear estimators $\hat{\nu}(x_t)$ for estimating $\nu(x_t)$ in the step 1 of the method outlined above.

The method of selecting the bandwidth in the local linear estimation is mentioned in the Section 2.3: We select the bandwidth h with a sequence from 0.05 to 1 by 0.05, and then separate the dataset into ten parts, calculate the cross-validation error of different bandwidth h we select, then choose the bandwidth h which have the minimum cross-validation error.

To test the estimation is good or not, we compare the results with the cases when we assume the knowledge of the volatility function σ_t in (2.10), and estimate the ϕ_1 and ϕ_2 in(2.11) by Yule-Walker estimation. Since the function is not available in reality, we call these estimation as oracle WLSE.

The following table shows the sampling means and standard deviations for the proposed

estimators $(\hat{\phi}_1, \hat{\phi}_2)$ and the oracle estimators $(\bar{\phi}_1, \bar{\phi}_2)$ based on the 1000 simulations with $T=2000$ data points.

The performance of the proposed estimators $(\hat{\phi}_1, \hat{\phi}_2)$ is close to the ϕ_1, ϕ_2 we assume, and according to the MSE, the estimators are consistency. But the estimators are still under performance compare to the oracle estimators.

What's more, we can find that with the change of variance function, the difference between the estimators $(\hat{\phi}_1, \hat{\phi}_2)$ with the oracle estimators $(\bar{\phi}_1, \bar{\phi}_2)$ would also change. In our simulation, we can find under the variance function $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$, the estimators $(\hat{\phi}_1, \hat{\phi}_2)$ are much closer to the oracle estimators than the other one.

In addition, the standard deviations of the estimators $(\hat{\phi}_1, \hat{\phi}_2)$ would also change with the change of variance function, and we can find that the one with $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ performance much better.

Because of these above points, we can find that though we can get good estimation results by the two-lag method, however the performance of it would be influence by the variance function we select. And the estimation results are not as good as the oracle estimators.

(ϕ_1, ϕ_2)	Mn(Std) $\hat{\phi}_1$	Mn(Std) $\hat{\phi}_2$	Mn(Std) $\bar{\phi}_1$	Mn(Std) $\bar{\phi}_2$	MSE($\hat{\sigma}$)
(0.4,0.3)	0.3916(0.04)	0.2956(0.03682)	0.3989(0.02124)	0.2976(0.02144)	0.002539
(0.3,0.6)	0.2974(0.04081)	0.5925(0.04977)	0.3(0.0182)	0.597(0.01751)	0.00255
(0.6,0.3)	0.5869(0.09265)	0.2993(0.08453)	0.5992(0.02078)	0.2984(0.02069)	0.003595
(0.6,0.0)	0.5881(0.04827)	0.006651(0.03611)	0.5986(0.01998)	-0.000128(0.01695)	0.003128
(0.6,-0.3)	0.5845(0.06191)	-0.2872(0.04814)	0.5987(0.02115)	-0.2994(0.02103)	0.00484
(-0.3,-0.6)	-0.2953(0.03644)	-0.5867(0.04624)	-0.2995(0.01779)	-0.5985(0.01813)	0.009885
(-0.6,-0.3)	-0.5892(0.05609)	-0.289(0.04687)	-0.6007(0.02219)	-0.3001(0.02067)	0.004757
(0.4,-0.6)	0.3944(0.03918)	-0.5892(0.04546)	0.3997(0.01811)	-0.6004(0.01841)	0.01041

Table 2.1
Sampling mean and standard deviations based on 1000 simulations with
T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$

(ϕ_1, ϕ_2)	Mn(Std) $\hat{\phi}_1$	Mn(Std) $\hat{\phi}_2$	Mn(Std) $\bar{\phi}_1$	Mn(Std) $\bar{\phi}_2$	MSE($\hat{\sigma}$)
(0.4,0.3)	0.3979(0.02162)	0.2990(0.02158)	0.398(0.02163)	0.2991(0.02159)	0.0006428
(0.3,0.6)	0.2985(0.01814)	0.5984(0.01799)	0.299(0.01816)	0.5972(0.01793)	0.000577
(0.6,0.3)	0.5986(0.0217)	0.2982(0.02109)	0.5988(0.02177)	0.2986(0.02107)	0.0006442
(0.6,0.0)	0.5873(0.02286)	0.00136(0.02258)	0.5975(0.02059)	0.00007541(0.01735)	0.0007063
(0.6,-0.3)	0.5995(0.02075)	-0.2991(0.02144)	0.5997(0.02067)	-0.3005(0.021144)	0.0008462
(-0.3,-0.6)	-0.2996(0.01841)	-0.597(0.01796)	-0.2997(0.01832)	-0.5983(0.01783)	0.001278
(-0.6,-0.3)	-0.5988(0.02134)	-0.2989(0.02174)	-0.5989(0.02126)	-0.2999(0.02159)	0.000888
(0.4,-0.6)	0.4003(0.01816)	-0.598(0.01816)	0.4001(0.01838)	-0.5991(0.01805)	0.001188

Table 2.2
Sampling mean and standard deviations based on 1000 simulations with
T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$

3. Estimation based on general quadratic forms

The first section in this part mentioned the basic idea and the algorithm of the estimation based on general quadratic forms, and in the second section we present our simulation based on this method.

3.1 Basic idea and algorithm

Figueroa-López(2013a) [2] mentioned a new method to estimate this kind of the model based on general quadratic forms. The following result is obtained in that paper.

For a quadratic form statistics

$$\eta_t^2 = \left(\sum_{i=0}^m a_i y_{t-i} \right)^2 = \sum_{i,j=0}^m a_i a_j y_{t-i} y_{t-j} \quad (3.1)$$

And a more general quadratic form

$$\psi_t = \sum_{i,j=0}^m c_{i,j} y_{t-i} y_{t-j} \quad (3.2)$$

As we did in the previous section, we first need to find conditions for

$$E\psi_t = \sigma^2 \quad (3.3)$$

in the $AR(1)$ model with constant variance function.

Proposition 3.1 Under the model

$$y_t = \sigma_t v_t, \sigma_t = \sigma(x_t), \quad (3.4)$$

$$v_t = \phi_1 v_{t-1} + \epsilon_t \quad (3.5)$$

based on AR(1) error with $\sigma_t \equiv \sigma$ and under a symmetric design matrix $C = [c_{i,j}]$, the following are necessary and sufficient conditions for (3.2) to satisfy (3.3):

$$\begin{aligned} (i) \quad & \sum_{j=0}^m c_{j,j}^2 = 1, \\ (ii) \quad & 1 + \sum_{j=0}^{m-2} c_{j,j+2} = 0, \\ (iii) \quad & \sum_{j=0}^{m-i} c_{j,j+i} = 0 \end{aligned}$$

Moreover, under the previous conditions (i)-(iii), the statistic ϕ_t is such that

$$E\psi_t = \frac{\sigma^2}{1 + \phi_2} \quad (3.6)$$

for the general model with (3.4)-(3.5) with $\sigma_t \equiv \sigma$.

To simplify the analysis, we then assume that the matrix $C = [c_{i,j}]$ has the decomposition:

$$C = \sum_{l=1}^k \lambda_l a_l a_l^T \quad (3.7)$$

where $a_l := [a_{0,l}, \dots, a_{m,l}]^T$ are suitable linearly independent vectors. If we take $k = 1, \lambda_1 = 1, a_1 = [a_0, \dots, a_m]^T$, we can get the quadratic form statistics (3.1). And the resulting quadratic form ψ will be:

$$\psi_t = \sum_{l=1}^k \lambda_l \eta_{t,l}^2 := \sum_{l=1}^k \lambda_l \left(\sum_{i=0}^m a_{i,l} y_{t-i} \right)^2 \quad (3.8)$$

The conditions (i)-(iii) under the structure (3.7) can be noted as

$$\begin{aligned} (i) \quad & \sum_{l=1}^k \sum_{j=0}^m a_{j,j}^2 = 1, \\ (ii) \quad & 1 + \sum_{l=1}^k \lambda_l \sum_{j=0}^{m-2} a_{j,l} a_{j,j+2} = 0, \\ (iii) \quad & \sum_{l=1}^k \lambda_l \sum_{j=0}^{m-i} a_{j,l} a_{j,j+i} = 0, \forall i \in \{1, 3, \dots, m\} \end{aligned}$$

Then consider the situation when $m = 2, k = 2$. The conditions (i)-(iii) can be written as

$$\begin{aligned}
(i) \quad & \sum_{l=1}^2 \sum_{j=0}^2 a_{j,j}^2 = 1, \\
(ii) \quad & 1 + \sum_{l=1}^2 \lambda_l a_{0,l} a_{2,l} = 0, \\
(iii) \quad & \sum_{l=1}^2 \lambda_l \{a_{0,l} a_{1,l} + a_{1,l} a_{2,l}\} = 0
\end{aligned}$$

From (i) and (ii), we could get

$$\sum_{l=1}^2 \{(a_{0,l} + a_{2,l})^2 + a_{1,l}^2\} = 0.$$

Because of this, it is clear that there exists two cases:

$$(a) \lambda_1, \lambda_2 > 0, (b) \lambda_1 > 0, \lambda_2 < 0.$$

For the first case, we can get $a_{0,l} = -a_{2,l}$, which is exactly the two-lag difference estimator(1.4). Then we come to consider the second case. We first assume without loss of generality that $\lambda_1 = -\lambda_2 = 1$, then we have,

$$\begin{aligned}
(i) \quad & \sum_{j=0}^2 (a_{j,1}^2 - a_{j,2}^2) = 1, \\
(ii) \quad & 1 + 2(a_{0,1} a_{2,1} - a_{0,2} a_{2,2}) = 0, \\
(iii) \quad & a_{1,1}(a_{0,1} + a_{2,1}) = a_{1,2}(a_{0,2} + a_{2,2})
\end{aligned}$$

After plug in conditions (i) and (ii) in (3.3), we can get that

$$\psi_t = \sum_{j=0}^2 (a_{j,1}^2 - a_{j,2}^2) y_{t-j}^2 - y_{t-1} y_{t-2} + 2(a_{0,1} a_{1,1} - a_{0,2} a_{1,2}) y_{t-1} (y_t - y_{t-2})$$

Then, we add a new condition

$$(iv) a_{0,1} + a_{1,1} + a_{2,1} = a_{0,2} + a_{1,2} + a_{2,2} = 0, \tag{3.9}$$

so that ψ_t will be the difference of the squares of two filters. Constraints (iii)-(iv) yield that $a_{1,1}^2 = a_{1,2}^2$ and ψ_t simplifies to

$$\psi_t = (a_{0,1}^2 - a_{0,2}^2)(y_t^2 - t_{t-2}^2) + y_{t-2}^2 - y_t y_{t-2} + 2(a_{0,1} \pm a_{0,2})a_{1,1}y_{t-1}(y_t - y_{t-2}),$$

which essentially is of the general form

$$\psi_t = a(y_t^2 - y_{t-2}^2) + y_{t-2}^2 - y_t y_{t-2} + 2by_{t-1}(y_t - y_{t-2}) \quad (3.10)$$

a and b are constants. So we can get a proposition that shown in Figueroa-López(2013a) [2]

Proposition 3.2 Let ψ_t be of the form

$$\psi_t = (a_{0,1}y_t + a_{1,1}y_{t-2} + a_{2,1}y_{t-2})^2 - (a_{0,2}y_t + a_{1,2}y_{t-2} + a_{2,2}y_{t-2})^2,$$

under the filter constraint (3.9). Then, $E\psi_t^2 \equiv \sigma^2$ under (3.4)-(3.5) with $\sigma_t \equiv \sigma$ and $\phi_2 = 0$ if and only if ψ_t is of the form (3.10) for reals a and b .

Combined Propositions 3.1 and 3.2 with the algorithm in Section 2, we can obtain the following method:

1. Estimate $\nu_t^2 := \nu^2(x_t) := \frac{\sigma^2(x_t)}{1+\phi_2}$ by

$$\tilde{\nu}_{t,1}^2 := \frac{1}{2}(\hat{\nu}_{t,1}^2 + \hat{\nu}_{t,2}^2),$$

where $\hat{\nu}_{t,1}^2 := \hat{\nu}_1^2(x_t)$ and $\hat{\nu}_{t,2}^2 := \hat{\nu}_2^2(x_t)$ are the local linear estimators of the following two non-parametric regressions:

$$\psi_{t,1}^2 := \nu_1^2(x_t) + \tilde{\epsilon}_{t,1}, \psi_{t,2}^2 := \nu_2^2(x_t) + \tilde{\epsilon}_{t,2}, t = 4, \dots, T,$$

with $\psi_{t,1}^2, \psi_{t,2}^2$ is the function defined in equation (3.10) with two different parameter values (a_1, b_1) and (a_2, b_2) .

2. Standardize the observations $\tilde{y}_t := \tilde{\nu}_t^{-1}y_t$ and then estimate (ϕ_1, ϕ_2) via the WLSE:

$$\tilde{\phi}_2 := (\tilde{A}^2 - \tilde{B}^2)^{-1}(\tilde{A}\tilde{C} - \tilde{B}^2), \tilde{\phi}_1 = \tilde{A}^{-1}\tilde{B}(1 - \tilde{\phi}_2) \quad (3.11)$$

with $\tilde{A} = \sum_{t=4}^T \tilde{y}_t^2, \tilde{B} := \sum_{t=4}^T \tilde{y}_t \tilde{y}_{t-1}, \tilde{C} := \sum_{t=4}^T \tilde{y}_t \tilde{y}_{t-2}$.

3. Estimate $\sigma_t^2 := \sigma^2(x_t)$ by

$$\tilde{\sigma}_t^2 := (1 + \hat{\phi}_2)\hat{\nu}_t^2, \quad or \quad \tilde{\sigma}_{t,1}^2 := (1 + \hat{\phi}_2)\hat{\nu}_{t,1}^2, \quad (3.12)$$

3.2 Simulation result

In this part, we run the simulation base on the algorithm mentioned above. The model and the dataset we used is the same as the one we used in Section 2.

We use two set of a_1, a_2, b_1, b_2 , i.e.

$$(i) a_1 = a_2 = \frac{1}{2}, b_1 = 0, b_2 = 30,$$

$$(ii) a_1 = a_2 = \frac{1}{2}, b_1 = -b_2 = 10$$

And with two set of variance function, i.e.

$$(a) \quad \sigma_t^2 = (x_t + 0.1)^2, \quad (b) \quad \sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$$

$\tilde{\phi}_{1,a}, \tilde{\phi}_{2,a}, MSE(\tilde{\sigma}_a)$ is the estimation of the case(i), $\tilde{\phi}_{1,b}, \tilde{\phi}_{2,b}, MSE(\tilde{\sigma}_b)$ is the estimation of the case(ii).

(ϕ_1, ϕ_2)	Mn(Std) $\tilde{\phi}_{1,a}$	Mn(Std) $\tilde{\phi}_{2,a}$	$MSE(\tilde{\sigma}_a)$
(0.4,0.3)	0.3983(0.02360)	0.3007(0.02302)	0.001631
(0.3,0.6)	0.2981(0.02019)	0.5995(0.02079)	0.001429
(0.6,0.3)	0.5966(0.02468)	0.3018(0.02451)	0.001637
(0.6,0.0)	0.5974(0.02456)	0.002256(0.0242)	0.01892
(0.6,-0.3)	0.5986(0.02362)	-0.2973(0.02266)	0.009689
(-0.3,-0.6)	-0.3007(0.02909)	-0.5955(0.02879)	0.008471
(-0.6,-0.3)	-0.5975(0.03073)	-0.2959(0.03064)	0.007132
(0.4,-0.6)	0.3982(0.03099)	-0.5955(0.0304)	0.008326

Table 3.1

Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$ and the parameter (i) $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$

(ϕ_1, ϕ_2)	Mn(Std) $\tilde{\phi}_{1,a}$	Mn(Std) $\tilde{\phi}_{2,a}$	$MSE(\tilde{\sigma}_a)$
(0.4,0.3)	0.398(0.0219)	0.2998(0.02159)	0.0005459
(0.3,0.6)	0.2991(0.0194)	0.5982(0.0185)	0.0004832
(0.6,0.3)	0.5983(0.02251)	0.2995(0.02217)	0.0005773
(0.6,0.0)	0.5991(0.02245)	0.0002004(0.02242)	0.0005773
(0.6,-0.3)	0.5985(0.02089)	-0.2989(0.02112)	0.0009743
(-0.3,-0.6)	-0.3002(0.01792)	-0.5979(0.01835)	0.0009974
(-0.6,-0.3)	-0.598(0.02187)	-0.299(0.02182)	0.0006684
(0.4,-0.6)	0.3988(0.01785)	-0.5972(0.01775)	0.0007159

Table 3.2

Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ and the parameter (i) $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$

(ϕ_1, ϕ_2)	Mn(Std) $\tilde{\phi}_{1,b}$	Mn(Std) $\tilde{\phi}_{2,b}$	$MSE(\tilde{\sigma}_a)$
(0.4,0.3)	0.3973(0.02359)	0.3013(0.02335)	0.002519
(0.3,0.6)	0.2976(0.01799)	0.5998(0.01815)	0.003423
(0.6,0.3)	0.5974(0.02234)	0.2998(0.02301)	0.003299
(0.6,0.0)	0.5981(0.0261)	0.001963(0.02694)	0.002847
(0.6,-0.3)	0.6002(0.02585)	-0.2986(0.02457)	0.003795
(-0.3,-0.6)	-0.2982(0.02704)	-0.5946(0.02614)	0.006762
(-0.6,-0.3)	-0.5976(0.02478)	-0.2959(0.02581)	0.003892
(0.4,-0.6)	0.3978(0.02616)	-0.5954(0.02609)	0.006681

Table 3.3

Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (a) $\sigma_t^2 = (x_t + 0.1)^2$ and the parameter (ii) $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$

(ϕ_1, ϕ_2)	Mn(Std) $\tilde{\phi}_{1,b}$	Mn(Std) $\tilde{\phi}_{2,b}$	$MSE(\tilde{\sigma}_a)$
(0.4,0.3)	0.3983(0.0204)	0.3004(0.02057)	0.0005232
(0.3,0.6)	0.2974(0.01837)	0.5989(0.01794)	0.0006572
(0.6,0.3)	0.5965(0.02121)	0.3011(0.02122)	0.0006568
(0.6,0.0)	0.5992(0.02271)	0.0002619(0.0217)	0.000587
(0.6,-0.3)	0.5976(0.02164)	-0.2985(0.02079)	0.0006479
(-0.3,-0.6)	-0.299(0.01811)	-0.5974(0.01812)	0.0008999
(-0.6,-0.3)	-0.6002(0.0207)	-0.2989(0.022173)	0.0006868
(0.4,-0.6)	0.3988(0.01853)	-0.5978(0.01759)	0.0008983

Table 3.4

Sampling mean and standard deviations based on 1000 simulations with T=2000 design points with variance function (b) $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ and the parameter (ii) $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$

Comparing these two tables with the one in section 2, we can find that this new estimation algorithm behaves much better than the first estimation method.

We also find some interesting facts about the relationship between the choice of parameter a, b in (3.10) with the estimation of the variance $\tilde{\sigma}^2$: When we use the first set of parameter $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$, we can find that the estimator $\tilde{\phi}_1, \tilde{\phi}_2$ performance well. However, when we consider $\tilde{\sigma}^2$ mentioned in (3.12), we can find that the one used $\hat{\nu}_{t,1}^2$ performs much better than the other one, which used the combination of $\hat{\nu}_{t,1}^2$ and $\hat{\nu}_{t,2}^2$. We choose the estimation of $\tilde{\sigma}^2$ based on the simulation of $\phi_1 = 0.3, \phi_2 = 0.6$, variance function $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$ shows below:

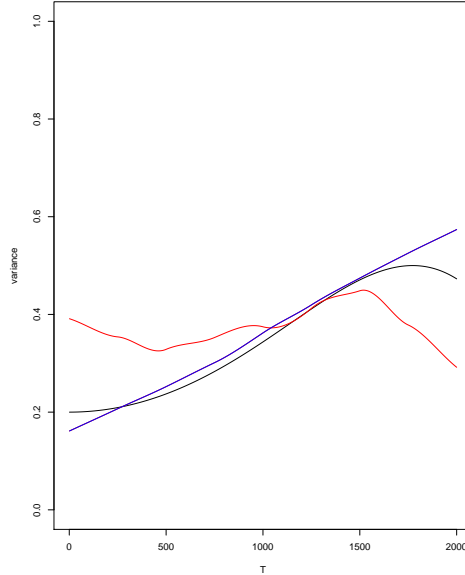


Figure 3.1. Estimation of variance where $a_1 = a_2 = 1/2, b_1 = 0, b_2 = 30$

In the graph above, the black line is the variance function $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$, the red line is $\tilde{\sigma}_t^2 := (1 + \hat{\phi}_2) \hat{\nu}_t^2$, the blue line is $\tilde{\sigma}_t^2 := (1 + \hat{\phi}_1) \hat{\nu}_t^2$. It is clearly that the blue line which used only $\hat{\nu}_{t,1}^2$ performs better. The reason to cause this problem maybe is that the parameter b_2 we choose here is too extreme.

However, when we consider the second set of parameter $a_1 = a_2 = 1/2, b_1 = 10, b_2 = -10$, for the same simulation, when we consider $\tilde{\sigma}^2$ mentioned in (3.12), we can find that the one used the combination of $\hat{\nu}_{t,1}^2$ and $\hat{\nu}_{t,2}^2$ performs much better than the other one. The estimation of $\tilde{\sigma}^2$ based on the same parameter and variance function as the former graph shows below:

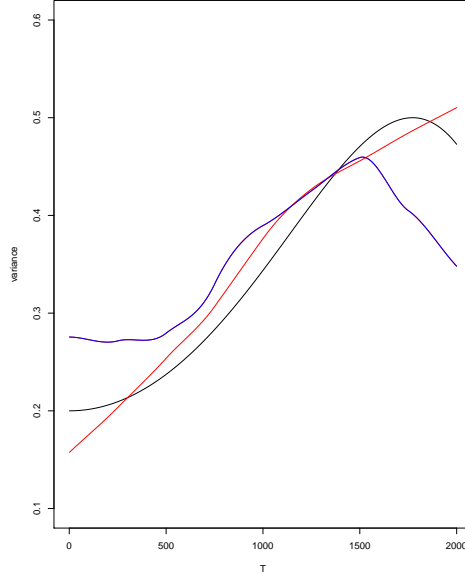


Figure 3.2. Estimation of variance where $a_1 = a_2 = 1/2, b_1 = -b_2 = 10$

In the graph above, the black line is the variance function $\sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$, the red line is $\tilde{\sigma}_t^2 := (1 + \hat{\phi}_2)\hat{\nu}_t^2$, the blue line is $\tilde{\sigma}_t^2 := (1 + \hat{\phi}_2)\hat{\nu}_t^2$. It is clearly that the red line which use the combination of $\hat{\nu}_{t,1}^2$ and $\hat{\nu}_{t,2}^2$ performs better.

Based on this discussion, we can find that the choice of $\tilde{\sigma}^2$ mentioned in (3.12) based on the parameter a_1, a_2, b_1, b_2 we choose. In our further study, we could focus on the choice of these parameters and find their relationship with the estimation consistency.

4. Conclusion

In this thesis we do the estimation of a non-parametric volatility model with autoregressive error of order two based on two-lag difference and general quadratic forms.

According to the simulation based on two different variance functions:

$$(a) \quad \sigma_t^2 = (x_t + 0.1)^2, \quad (b) \quad \sigma_t^2 = 0.3 \sin(2x_t^2) + 0.2$$

on 1000 simulations with $T=2000$ data points, we can find that both methods can do a good estimation for the parameters ϕ_1 and ϕ_2 , and the small MSE show that the estimation is consistence.

Comparing this two method we can find that the one based on general quadratic forms can get a closer estimation for the parameters, and sometimes it could even do a better estimation than the oracle estimators.

For our further study, we can try to find some more details about the choice of parameter a, b in (3.10) for the general quadratic forms method, to help the improvement of the performance of the estimation consistency.

REFERENCES

- [1] José E Figueroa-López and Michael Levine. Nonparametric regression with rescaled time series errors. *Journal of Time Series Analysis*, 34(3):345–361, 2013.
- [2] José E Figueroa-López. Nonparametric estimation of time series volatility model with general quadratic terms. *unpublished manuscript*, 2013.
- [3] Christian M Dahl and Michael Levine. Nonparametric estimation of volatility models with serially dependent innovations. *Statistics & probability letters*, 76(18):2007–2016, 2006.
- [4] Tsung-Chih Lin, Mohsen Pourahmadi, and Anton Schick. Regression models with time series errors. *Journal of Time Series Analysis*, 20(4):425–433, 1999.
- [5] Peter Hall and Ingrid Van Keilegom. Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):443–456, 2003.
- [6] Qin Shao and LJ Yang. Autoregressive coefficient estimation in nonparametric analysis. *Journal of time series analysis*, 32(6):587–597, 2011.
- [7] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112, pages 181–183. Springer, 2013.
- [8] B. D. Ripley. *Local Polynomial Regression Fitting*. R Documentation.