

Forecasting Economic Variables Using Markov Quantile Regression Approach

Xiaoxiao Tang

B.S., Tsinghua University, P.R.China, 2009

A Dissertation Presented to the Graduate Faculty
of University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Statistics

University of Virginia
May, 2014



Abstract

The classical least squares estimation approach has several potential problems when applied to economic variables. Quantile regression, on the other hand, provides us with an alternative way to work in non-Gaussian settings and obtain a more complete picture on the distribution of the forecasted variables. However, how to utilize this extra distributional information to obtain an optimized forecast is of great concern. This thesis intends to overcome this problem and explore the application of quantile regression in the examination of relationship among economic variables and in the prediction of equity returns.

We develop a new methodology that incorporates market evolvment dynamics to combine quantile regression results. We assume that the market state movement is controlled by a three state Markov chain, and the expected return at each state is proportional to different quantiles. The prediction is given as a weighted sum of estimated quantiles where the weights are the transition probabilities. By including market movement process in the prediction, we can obtain a way to combine different quantiles into an optimized point forecast. The empirical results show that our model improves both in sample fitness and prediction accuracy.

Our first application of the new specification is to forecast equity risk premium. With the application of the combination method, we show that our model outperforms both OLS and fixed weighted quantile regression model. We also employ the model to check the relationship between excess return of portfolios and the three Fama French factors. The proposed model fits the data better than OLS, and has greater predictive power. The future work includes generalization of the model and issues regarding portfolio construction.

Acknowledgements

I would like to give my deepest appreciation to all the people who supported and encouraged me over the past five years.

I cannot express enough thanks to Professor Feifang Hu, who has been a mentor, colleague and friend. His enduring guidance and advice lightened my way. From the weekly meeting with him, not only did I learn about the techniques that solve specific problems, but I also acquire the methodology that will benefit me for the life time. His consistent support and precious advice are great encouragement to me when I face hardships.

Many thanks to Professor Michael Gallmeyer, who has shown the attitude and the substance of a genius. He started to help me from the most basic concept when I knew nothing about finance. He taught me how to read papers and generate interesting research topics. He also encouraged me to attend finance seminars, which were really valuable for me to get familiar with the field.

I am especially grateful to Professor Peiming Wang, who showed extreme patience in the face of numerous obstacles. Whenever there were tough problems in my research, he was always there to give me prompt guidance, which is strikingly treasurable due to our time difference and the fact that we mostly communicate by emails. His constructive ideas highly edified me when I built the model. Without his supervision and constant help, this dissertation would not have been possible.

In addition, a thank you to my committee members, Professor Daniel Keenan and Professor Jianhui Zhou. Thank you for all the insightful suggestions about my research and my career concerns. I also want to thank Professor Tingting Zhang, Professor Jeff Holt and Professor Dan Spitzner, whose passion for statistics has inspired me a lot. Thanks also goes to my fellow students in the department. Their

vitality and enthusiasm make Halsey Hall one of the most beautiful memories in my life.

Last but not least, thanks to my parents. I would end up nowhere without your love.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
2 The Markov Quantile Regression Approach	6
2.1 Quantile Regression Models	6
2.2 Our Specification	9
2.3 Predictive Methods	13
2.4 Evaluation Criteria	15
2.4.1 In Sample Evaluation	15
2.4.2 Out of Sample Evaluation	16
3 Theoretical Results	24
3.1 Autocorrelation Structure	24
3.2 Quantile Regression Models	26
3.3 Markov Chain	33
4 Numerical Study	47
5 Empirical Results	67

5.1	Equity Risk Premium Forecast	67
5.1.1	Data Description	67
5.1.2	Predictive Model	69
5.1.3	Empirical Results	74
5.1.4	Robustness Check	79
5.2	Fama-French Three Factor Model	83
5.2.1	Data Description	83
5.2.2	Fama-French Three Factor Model	83
5.2.3	Difference Between Two models	88
5.2.4	In Sample Fitness	90
5.2.5	One time period lag model	92
6	Conclusions and Possible Future Directions	97
	Bibliography	101

Chapter 1

Introduction

The research of the prediction of portfolio return and equity risk premium is intensive, since it is of great value to both academics and practitioners. Different variables and different techniques are employed in the related literature. Most of the previous research are relied on the classical least-squares method and maximum likelihood estimators. However, there are potential problems with these approaches. First, it requires a Gaussian assumption for the return distribution. We know that there are numerous evidence showing the asymmetric leptokurtic feature of stock return distributions. Therefore the Gaussian assumption is questionable. Second, these methodologies only offer a conditional mean view of the relationship between the predictors and the forecasted variable, which indicates that the covariates are restricted to only contribute to a location shift of return distributions. However, the predictors are allowed to influence the whole conditional distributions. Quantile regression method offers us a broader view and a more complete picture of this relationship, so that enhances our possibilities to more clearly understand the relationship among economic variables, and ultimately, to a better forecast.

My PhD thesis aims to explore the application of quantile regression in the field

of finance. Koenker and Bassett (1978) propose the quantile regression methodology, which gives us a brand new way to work in non-Gaussian settings, as well as the possibility to examine the behavior of the whole conditional distribution of the response variables. Both theoretical and empirical research have been carried out in the related area, including model extensions, new inferential procedures and various empirical applications. However, the application of quantile regression in the area of finance has not been explored until quite recently. Most of the work has focused on risk management (VaR model). Other applications include asset management (Feng, Chen and Basset (2008), Ma and Pohlman (2008)), the analysis of the cross section of stock market returns (Barnes and Hughes (2002)) and hedge fund pricing (Meligkotsidou, Vrontos and Vrontos (2009)).

With quantile regression, we are able to obtain a clearer idea of the conditional distribution. The question is how to use this information to give better prediction. To address this problem, we take the market condition into consideration. We assume that there are three states in the stock market: bull, normal and bear. Given the state, there is a conditional distribution to model the stock return behavior. We use a Markov chain to model the market evolution dynamics, and use quantile regression results to estimate the forecasted variable. The mean of the forecasted variable is given as a combination of the three forecasts at different states, and the weights are determined by the state of the previous time period. This is our main idea of the application of quantile regression.

To legalize the use of our model, we present the theoretical results. Firstly, we make assumptions on the autocorrelation structure of the time series data. Under AR model set up and some mild assumptions about residuals, we show the asymptotic properties of the estimation of quantile regression. Moving forward, we also prove the consistency of the estimation of the parameters related with Markov chain,

including the state variables and transition probabilities. What is more, we obtain the asymptotic normality of the transition probabilities estimates.

Unfortunately, the asymptotic distribution of the coefficient estimation in quantile regression model is irregular, which means that we do not have the normality to support statistical inference. To overcome this difficulty, we perform simulation study on the time series data with quantile regression model. The simulation results confirm our consistency property. We also find that the asymptotic distribution is close to normal. Such results justify our use of confidence interval in estimation. Then we use two sets of data to do the empirical analysis.

The first set of data is the quarterly data from Welch and Gayol (2008). With this set of data, we aim to explore the application of quantile regression in equity risk premium forecast. The equity risk premium is calculated as the difference between the continuously compounded S&P 500 returns and the Treasury Bill rate. We select 15 predictors, consistent with literature, to do the regression. The sample spans from 1947:1 to 2011:4. We split it into in-sample and out-of-sample period. The in-sample period data is used to do the model estimation, and the equity risk premium for out-of-sample period is predicted by an expanding window. We apply the combination methods proposed by Rapach, Strauss and Zhou (2009) to obtain more stable and reliable results. Besides the widely used benchmark, historical average, OLS prediction is also considered as a baseline model. We compare our method with the fixed weight quantile regression approach proposed by Meligkotsidou et. al. (2012) as well. Their method is a simple combination of different quantiles given by quantile regression model, so they do not utilize the market evolution information. We make this comparison to test whether such information contributes significantly to the forecast. We choose three model comparison criteria to compare the models. R_{os}^2 tests the out of sample performance in the sense of mean squared prediction

error; $ENC - T$ tells whether one model contains useful information for prediction that is not present in another competing model; ΔU , the utility difference, and ΔW , the wealth gain, give the economic evaluations of the models. Results show that our model significantly improves the forecast from benchmark models, and the market movement information is crucial to precise prediction.

The other data set we use is monthly return data of thirty industry portfolios. The data is obtained from CRSP database. We use this data set to explore the use of quantile regression in portfolio construction. OLS estimates of Fama French three factor model is used as the benchmark. We again apply the quantile regression combined with a Markov chain to do the prediction, with much shorter tails as the extreme states. The estimation is conducted for both models where the predictors and portfolio returns are in the same time period and where the latter is one time period ahead of the former. We do in sample analysis for both cases. We apply mean squared error, and the less commonly used mean absolute deviation as model comparison criteria. The first reason that we consider the absolute value of the prediction errors is due to the possibility of heavy tail distributions of portfolio returns. In this case, the usual mean squared error puts too much weight on the extreme values. The second reason is that mean squared error is served as the loss function of OLS method. Therefore we apply a more robust criterion to do the test so that we can get a clearer view. We also carry out the out of sample analysis for one month lag model, since the one period lag structure resembles the real world when an investor makes prediction. We use a fixed window to do the prediction. Results show that our model outperforms OLS for both in sample and out of sample case. The application of quantile regression to portfolio construction is left for future study.

The rest of the thesis is organized as follows. Chapter 2 describes the model.

Theoretical results are presented in Chapter 3. Numerical study is contained in Chapter 4. Chapter 5 gives the empirical results and Chapter 6 concludes the project.

Chapter 2

The Markov Quantile Regression Approach

2.1 Quantile Regression Models

Assume Y is the variable whose future values are of interest, such as equity excess return. $X = \{x_1, x_2, \dots, x_p\}$ is a set of predictors that are assumed to influence Y . The simplest way to model the relationship between Y and X is a linear regression model:

$$Y = X^T \beta + \epsilon.$$

Here, we assume each predictor in X is linearly correlated with the response Y . The error term ϵ is assumed to be normally distributed with mean 0 and a constant standard deviation σ . The coefficient vector β is interpreted as the partial contribution of the factors X to the mean of the response Y . Under these assumptions, given values of X , the conditional distribution of Y is normal, expressed as,

$$Y | X \sim N(X^T \beta, \sigma^2).$$

Now consider a data set of n observations, $\{(y_1, X_1), (y_2, X_2), \dots, (y_n, X_n)\}$, The estimates of β are obtained by minimizing the square loss function

$$\hat{\beta} = \arg \min_{b \in B} \sum_{i=1}^n (y_i - x_i^T b)^2,$$

where B is the parameter space. The loss function of OLS method aims to minimize the sum of squared deviations between individual observation, y_i , and its fitted value. Therefore, the in-sample prediction given by OLS regression is an estimate of the conditional mean of the response variable. Given $X = X_0$, the predicted value of Y is obtained by:

$$\hat{Y} = X_0^T \hat{\beta}.$$

However, this standard regression approach may not be appropriate in cases where the distribution of Y exhibits some particular characteristics, for instance, a high degree of non-normality, fat tails, excess kurtosis and skewness. For example, in the case of fat tails, the square loss function puts too much weight on the extreme values. Another example is when there is extreme skewness or outliers, mean prediction is not as robust as forecasts based on quantiles. In the presence of these characteristics the conditional mean specification for the response Y may not capture the effect of the independent variables to the entire distribution of Y and may provide estimates which are not robust.

On the other hand, the quantile linear regression approach provides an alternative to standard linear regression for examining the impact of independent variables on the entire distribution of Y . Quantile regression is first introduced by Koenker and Bassett (1978) and has increasingly attracted attention as a robust alternative in the field of economics and finance. This approach can be useful in uncovering potential

differences in the effect of the independent variables across quantiles of Y . A τ th quantile linear regression can be specified as follows:

$$Q_{Y|X}(\tau) = X^T\beta_\tau + \epsilon_\tau, \quad (2.1)$$

where $Q_{Y|X}(\tau)$ is the τ th quantile of the conditional distribution of Y given X . β_τ are interpreted as the partial effect of X to the τ th quantile of Y . ϵ_τ are assumed to have zero as its τ th quantile. Here in this model, we assume that X has a linear relationship with the τ th quantile of Y , and we do not put distribution assumption on the error term, ϵ . The estimations of β_τ , denoted as b_τ , are calculated by minimizing the following term:

$$b_\tau = \arg \min_{b \in B} \sum_{i=1}^n \rho_\tau(y_i - x_i^T b), \quad (2.2)$$

where $\rho(\cdot)$ is the check function defined by Koenker and Bassett (1978)

$$\rho_\tau(e) = (\tau - I(e < 0))e,$$

$$\text{where } I(e < 0) = \begin{cases} 1, & \text{when } e < 0 \\ 0, & \text{when } e \geq 0 \end{cases},$$

in which e is the residual and calculated by $e_i = y_i - x_i^T b$. Based on the loss function, we can see that we look for the values of the coefficients that minimize a weighted sum of absolute differences between the observations and $x_i^T b$. This provides an estimate of the τ th quantile of the conditional distribution of the response variable Y . Based on the estimation for different τ , we can estimate the conditional quantile function, which is the inverse of the conditional probability distribution function. Therefore quantile regression allows us for an estimate of the whole picture of the

distribution of the response variable, instead of focusing only on the conditional mean as OLS method does.

Now that we have two different tools for a linear model, which one should we apply? Since the basic difference between the two set-ups is that in the quantile regression, the partial effect of X is assumed to be different for different quantiles of Y , therefore the results of a tentative QR might be helpful. If the estimations b_τ vary significantly according to different quantiles, then we think QR will be more appropriate.

2.2 Our Specification

Assume $\{(y_t, X_t); t = 1, \dots, T\}$ is a time series data set where y_t is an economic variable whose future values are of interested and X_t is a vector of pre-determined values of explanatory variables which contain the information we have up to time T . This means that X_t can either take values of current time period when we are interested in the relationship between Y and X , or denote time lag values when we do prediction. Observation at time t , y_t , lies in one of the three ranges of the conditional distribution of the dependent variable, defined as $(-\infty, Q_{Y_t|X_t}(\tau_1))$, $[Q_{Y_t|X_t}(\tau_1), Q_{Y_t|X_t}(\tau_3))$ and $[Q_{Y_t|X_t}(\tau_3), +\infty)$, where $Q_{Y_t|X_t}(\tau_1)$ and $Q_{Y_t|X_t}(\tau_3)$ are low and high quantiles of the conditional distribution of Y_t given X_t , namely, $\tau_1 < 0.5$ and $\tau_3 > 0.5$. Let S_t be the state variable at time t , which takes three possible values $\{1, 2, 3\}$, each representing one of the three states of the stock market: bull, normal and bear, respectively. We use S_t as an indicator representing the range where y_t is. To model the serial dependency of observations over time, we assume that S_t follows a 3-state first-order Markov chain with the transition probabilities defined as

$$P(S_{t+1} = j \mid S_t = i) = P_{ij}, \quad i, j = 1, 2, 3. \quad (2.3)$$

Note that it is not hidden Markov chain because S_t is observable in the definition. Moreover, conditional on $S_t = 1$, it is assumed that Y_t does not exceed a low percentile $Q_{Y_t|X_t}(\tau_1)$ of the marginal distribution of asset returns. That is $P(Y_t > Q_{Y_t|X_t}(\tau_1)) = 0$ and $P(Y_t < Q_{Y_t|X_t}(\tau_1)) = 1$. Furthermore, we assume that conditional on S_t and X_t , the conditional mean and standard deviation are determined by

$$E(Y_t | S_t = 1) = c_1 \times Q_{Y_t|X_t}(\tau_1),$$

$$\sigma(Y_t | S_t = 1) = \sigma_1,$$

where c_1 is a constant, while the value of $Q_{Y_t|X_t}(\tau_1)$ depends on X_t . Statistically, the above assumptions about the first two moments are sensible because as defined above, the distribution conditional on $S_t = 1$ is a Pareto distribution, which has the property that its mean is proportional to the upper limit with a constant.

Similarly, conditional on $S_t = 3$, it is assumed that Y_t is larger than a high percentile $Q_{Y_t|X_t}(\tau_3)$ of the marginal distribution of asset returns, and the conditional mean and standard deviation are determined by

$$E(Y_t | S_t = 3) = c_3 \times Q_{Y_t|X_t}(\tau_3),$$

$$\sigma(Y_t | S_t = 3) = \sigma_3.$$

Conditional on $S_t = 2$, it is assumed that Y_t can only take a value between the two quantiles with the first two moments given by

$$E(Y_t | S_t = 2) = c_2 \times Q_{Y_t|X_t}(\tau_2),$$

$$\sigma(Y_t | S_t = 2) = \sigma_2,$$

where $Q_{Y_t|X_t}(\tau_2)$ is the median of the conditional distribution of y_t given X_t . by definition we assume $Q_{Y_t|X_t}(\tau_1) < Q_{Y_t|X_t}(\tau_2) < Q_{Y_t|X_t}(\tau_3)$.

By Baye's Theorem, we can get the property of the marginal distribution under this set up. Given X_t , the conditional mean of Y_t is given by

$$E(y_t|X_t) = P(S_t = 1)c_1Q_{Y_t|X_t}(\tau_1) + P(S_t = 2)c_2Q_{Y_t|X_t}(\tau_2) + P(S_t = 3)c_3Q_{Y_t|X_t}(\tau_3). \quad (2.4)$$

Note that the above model specification does not require the specification of the entire probability distribution for Y_t . By assigning different values of quantiles, we can examine differently types of distribution. For example, we can choose relatively extreme high and low quantiles for fat tail distribution; we can select asymmetric quantiles when there is excess kurtosis and skewness. Also, we are free to examine the tail behaviors for some specific quantiles. Our model can also be generalized to multi-state cases if more than three quantiles are of interest.

The new idea that lies into our specification is that we link the regression with the market evolution dynamics. We know that data obtained from the stock market are mostly time series data, which are hardly *i.i.d.*. On the contrary, there usually exhibits a high degree of autocorrelation. Such autocorrelation provides us with a lot of information about how the market changes, thus allows us for a deeper understanding of the behavior of the stock distribution.

Previous models, such as QRAD location and QRAD probability proposed by L. Ma and L. Pohlman (2008) has already presented some idea of linking the regression with market evolution, however, the dynamic is too simple. Our model is a generalization of both of the QRAD methods. The assumption of QRAD location is that stock returns remain the same quartile over periods. That is to say, they specify the

probability transition matrix as the identity matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The problem with this method is that it's very likely in reality that the stock returns sometimes could change dramatically and switch between different quartiles over times as we have seen in the stock markets recently. Alternatively, QRAD probability method assume the expectation of the marginal distribution takes the weighted average of forecasted quartiles.

$$E(y_t|X_t) = w_1Q_{Y_t|X_t}(\tau_1) + w_2Q_{Y_t|X_t}(\tau_2) + w_3Q_{Y_t|X_t}(\tau_3),$$

where w_i are a set of pre-assigned values and satisfy $\sum_{i=1}^k w_i = 1$. For example, if we choose $k = 3$ and $\{\tau_1, \tau_2, \tau_3\} = \{0.1, 0.5, 0.9\}$, then one reasonable selection of the weights could be $\{w_1, w_2, w_3\} = \{0.1, 0.8, 0.1\}$. This model is also one of the sub-models of our specification. They are equivalent if the probability transition matrix are chosen as

$$\begin{pmatrix} 0.1 & 0.8 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.8 & 0.1 \end{pmatrix}$$

Without the change of the conditional probabilities, this method does not take the evolution of the market into consideration. Another practical question is that by simply combining the predictions of different quartiles, what we get are predictions close to the mean, thus the improvement from OLS methods is unsatisfactory.

The main problem of QRAD methods lies with the *i.i.d.* assumption for returns.

To overcome this shortcoming of QR, It's worth trying to build some serial dependence structure in return forecast. Therefore we propose the above new specification that combines quantile regression with a Markov chain which models the evolution dynamic of the market.

2.3 Predictive Methods

One important issue in the study of equity risk premium is to predict future values. Quantile regression enables us to obtain the behavior of stock returns in different quantiles, the next question is, how can we take advantage of that information to better predict future stock returns?

For prediction of y_t , we consider the following:

$$\hat{y}_t = \hat{P}_{1S_{t-1}} \hat{c}_1 \hat{Q}_{Y_t|X_t}(\tau_1) + \hat{P}_{2S_{t-1}} \hat{c}_2 \hat{Q}_{Y_t|X_t}(\tau_2) + \hat{P}_{3S_{t-1}} \hat{c}_3 \hat{Q}_{Y_t|X_t}(\tau_3)$$

where $\hat{Q}_{Y_t|X_t}(\tau_k)$ is the estimated τ_k th conditional quantile regression given by

$$\hat{Q}_{Y_t|X_t}(\tau_k) = X_t^T \hat{\beta}_{\tau_k},$$

and \hat{c}_k is estimated as follows.

For each in-sample time period, we identify its market state by the comparison between the observation and the fitted values of each quantile.

$$\hat{S}_t = \begin{cases} 1, & \text{when } y_t < \hat{Q}_{Y_t|X_t}(\tau_1) \\ 3, & \text{when } y_t \geq \hat{Q}_{Y_t|X_t}(\tau_3) \\ 2, & \text{otherwise} \end{cases}$$

Based on the estimated values of S_t , we can identify each period with a state, and obtain three sub-samples, namely SS_1 , SS_2 and SS_3 , corresponding to each of the three states respectively. Then we use the sample standard deviations of the sub-samples as the estimates of the conditional standard deviations, σ_1 , σ_2 , and σ_3 respectively. As for the constants c_1 , c_2 and c_3 , we obtain their estimates by dividing the sample mean of a sub-sample by the average of the fitted quantile values for that sub-sample, *i.e.*,

$$\hat{c}_k = \frac{\bar{y}_{SS_k}}{\hat{y}_{\tau_k, SS_k}}. \quad (2.5)$$

It can be shown that this formula gives a consistent estimate of c_k .

The transition probabilities \hat{P}_{ij} are estimated in the statistical standard way. Denote

$$n_{ij} = \sum_{t=1}^{T-1} I(y_t \in SS_i \text{ and } y_{t+1} \in SS_j)$$

and

$$n_i = \sum_{t=1}^{T-1} I(y_t \in SS_i),$$

where again, $I(\cdot)$ is the indicator function. The estimate of the transition probability is thus given by:

$$\hat{P}_{ij} = \frac{n_{ij}}{n_i}. \quad (2.6)$$

As n_{ij} follows a Binomial distribution, the standard deviation of the transition probability is estimates by:

$$\hat{\sigma}_{P_{ij}} = \sqrt{\frac{\hat{P}_{ij}(1 - \hat{P}_{ij})}{T - 1}}.$$

Therefore, based on the above specification, we carry out the parameter estimation

as follows:

- 1. Fit return data to QR with three quantiles and identify each in sample period with one of the three market states S_t .
- 2. Assuming that $\{S_1, \dots, S_T\}$ follows a three-state first-order Markov chain, estimate the transition probabilities.
- 3. Based on the values of S_t , we divide the data set into three sub-samples, and estimate c_1, c_2 and c_3 .
- 4. Use the above prediction formula to give forecasts.

Notice that if we use a rolling window to do the predictions, all the parameter estimates are updated each period, including P_{ij} and c_i .

2.4 Evaluation Criteria

To measure the model performance, we introduce several evaluation criteria. Some of the criteria put more emphasis on the in sample performance, while others are more associated with out of sample performance, which is to say, they measure the predictive power of the model.

2.4.1 In Sample Evaluation

We consider two statistics to measure model's in sample performance. The first one is the widely used mean squared error, denoted as MSE , which is given by

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2.7)$$

where n is the sample size, y_i are the observations and \hat{y}_i are the fitted values given by the model. By calculating the mean of the squared deviations, MSE gives us a measure of the distance between the model and the real data.

In the basic OLS settings, the deviations are assumed to be normal, thus their distribution has relatively thin tails. However, in the non-normal settings, the distribution of the error term might exhibit some particular characteristics, like fat tails. In this case, the squared deviations put too much weight on the extreme values. Our model is distribution free, therefore we also consider another more robust measure, mean absolute deviation, denoted as MAD

$$MAD = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (2.8)$$

MAD is the average of the absolute deviations from the model to the data, therefore, it assigns each observation with the same weight, thus gives a more robust measure.

2.4.2 Out of Sample Evaluation

Out Sample R -Square

We use R_{os}^2 , proposed by Campbell and Thompson (2008) to compare the out of sample performance of two models. R_{os}^2 , as a comparison of model A and the standard benchmark model, historical average, is given by

$$R_{os}^2 = 1 - \frac{\sum_{s=1}^k (y_{m+s} - \hat{y}_{m+s,A})^2}{\sum_{s=1}^k (y_{m+s} - \bar{y}_{m+s})^2}, \quad (2.9)$$

where m is the number of time periods that are used for model estimation, and k is the length of the prediction period. $\hat{y}_{m+s,A}$ is response variable y at time $m + s$ forecasted by model A and \bar{y}_{m+s} is the average of the observations from the

beginning of the sample until time $m + s - 1$. This statistic measures the reduction in mean squared predictive error for model A relative to the natural benchmark model, historical average. When $R_{os}^2 > 0$, model A outperforms the benchmark.

Since the value of R_{os}^2 is sample dependent, we need to perform statistical tests to collect statistical evidence. The test we apply is the prediction accuracy test, or DM test proposed by Diebold and Mariano (1995). For notation convenience, define prediction error of model i as

$$u_{m+s,i} = y_{m+s} - \hat{y}_{m+s,i}$$

for time period $m + s$. Allow the prediction loss function to be an arbitrary function of the prediction error, $g(u_{m+s})$. The null hypothesis of DM test is that $E[g(u_{t,1})] = E[g(u_{t,2})]$ or $E[d_t] = 0$, where $d_t = g(u_{t,1}) - g(u_{t,2})$ is the loss difference. Thus, the "equal accuracy" null hypothesis is equivalent to the null hypothesis that the population mean of the loss differential series is 0. Consider $\{d_{m+s}\}_{s=1}^k$ as a sample of loss differences. We have

$$\sqrt{k}(\bar{d} - \mu) \xrightarrow{d} N(0, 2\pi f_d(0)), \quad (2.10)$$

where

$$\bar{d} = \frac{1}{k} \sum_{s=1}^k (g(u_{m+s,1}) - g(u_{m+s,2}))$$

is the sample mean of loss difference, and

$$f_d(0) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau)$$

is the spectral density of the loss difference at frequency zero, and

$$\gamma_d(\tau) = E((d_t - \mu)(d_{t-\tau} - \mu))$$

is the autocovariance of the loss difference at displacement τ , and μ is the population mean of loss difference. The formula for $f_d(0)$ shows that the correction for serial correlation can be substantial, even if the loss differential is only weakly serially correlated, due to cumulation of the autocovariance terms.

Specifically, in our case, the prediction is one time period ahead, which means that the prediction errors are independent. Therefore, the test reduces to the classical population mean test. We present the complete test above for future use. Here, we would like to apply DM test to test null hypothesis $R_{os}^2 = 0$ against the alternative hypotheses $R_{os}^2 > 0$. Define mean square prediction error of model A and historical average, as

$$MSPE_A = E(y_t - \hat{y}_{t,A})^2, MSPE_{bm} = E(y_t - \bar{y}_t)^2.$$

This pair of hypothesis is equivalent with $MSPE_A = MSPE_{bm}$ against $MSPE_A < MSPE_{bm}$. We choose the loss function $g(u_t) = u_t^2$, then we see that DM test can be applied to the test of R_{os}^2 . In this case, we have

$$d_{m+s} = u_{m+s,A}^2 - u_{m+s,bm}^2$$

and the test statistic is calculated as

$$DM = \sqrt{k} \frac{\bar{d}}{\hat{\sigma}_d}, \quad (2.11)$$

where \bar{d} and $\hat{\sigma}_d$ are the sample mean and sample standard deviation of d , respectively. The test statistic follows a t distribution, which in large sample case, can be approximated by a standard normal distribution.

ENC – T Test

The other test we apply is *ENC – T*, developed by A. Harvey, Leybourne and Newbold (1998). This test compares between two competing models, and checks whether one of them contains useful information for prediction that the other model does not. Consider a composite forecast $\hat{y}_{t,o}$, which is a convex combination of the forecast from model A, usually but not restricted to be the parsimonious model, $\hat{y}_{t,A}$ and another model B, $\hat{y}_{t,B}$. The optimal combination is assumed to be $\hat{y}_{t,o} = (1 - \lambda)\hat{y}_{t,A} + \lambda\hat{y}_{t,B}$, $0 \leq \lambda \leq 1$. $\lambda = 0$ means that the optimal forecast assigns zero weight to model B, which is defined as model A "encompasses" model B. The test based on *ENC – T* statistics is to test the null hypothesis, $\lambda = 0$, against the one-sided alternative hypothesis, $\lambda > 0$. The case of the null hypothesis is defined as model A is "conditionally efficient" with respect to model B by Granger and Newbold. Under the alternative hypothesis, model B contributes to the optimal prediction, therefore, contains information for forecast that is not present in model A.

Define the forecast errors of model A and B, respectively, as $u_{t,A} = y_t - \hat{y}_{t,A}$, $u_{t,B} = y_t - \hat{y}_{t,B}$, by the combination relationship among the three models, we can write:

$$u_{t,A} = \lambda(u_{t,A} - u_{t,B}) + u_{t,o}.$$

The combined forecast will then have smaller expected square error than model A unless the covariance between $u_{t,A}$ and $(u_{t,A} - u_{t,B})$ is 0. Given a series of forecast

errors, $(u_{t,A}, u_{t,B})$, $t = 1, 2, \dots, n$, to implement the test, it is natural to regress the differences on $u_{t,A}$ by OLS method to get the estimate and perform the standard regression-based test. However, Harvey, Leybourne and Newbold (1998) show that this widely used test statistics is not robust to non-normality. They propose the $ENC - T$ test to avoid this problem. Let

$$d_t = (u_{t,A} - u_{t,B})u_{t,A},$$

then the $ENC - T$ statistic is given by

$$ENC - T = \sqrt{k} \frac{\bar{d}}{\sqrt{\hat{Var}(d)}}, \quad (2.12)$$

where \bar{d} , $\hat{Var}(d)$ are the sample mean and sample variance of d over the out-of-sample period. Asymptotically, $ENC - T$ statistic follows a standard normal distribution under the null hypothesis. In the finite sample case, Harvey, Leybourne and Newbold (1998) show that we can apply the student's t distribution with $T - k - 1$ degrees of freedom. As the argument above, this is an upper tailed test. In practice, the natural benchmark(model A) is historical average. We also compare between the quantile regression models and OLS model.

The significance of $ENC - T$ test shows that model A contains useful information for forecast that is not captured by model B, however, this does not necessarily lead to the conclusion that model A outperforms model B. Rapach, Strauss and Zhou (2008) apply the test to individual models to collect statistical evidence for their combination method. Similar with what they do, to compare the performances between two models, we also run the test at the opposite direction. The other way

test is conducted by defining

$$d'_t = (u_{t,B} - u_{t,A})u_{t,B}$$

and calculating the statistics as the above equation. This test is equivalent with testing $\lambda = 1$ against $\lambda < 1$ in the above set-up. If both of the tests are significant, then we have the conclusion that $0 < \lambda < 1$, which means we should combine model A and model B to get the optimal forecast. However, if the test of model A against model B is significant and the other one is not, we can draw the conclusion that model A outperforms model B.

Economic Evaluations

At last, we also employ economic evaluation measures. The reason is that there are several limitations of statistical evaluation. One is that even if the test is significant under statistical tests, the value is still unreasonably small. For example, typical significant R_{os}^2 values are around 2 percent. However, a reduction of two percent of mean square prediction error does not mean too much to the market. The other limitation would be that the statistics lack economic intuition behind them. From Campbell and Thompson (2008) and Rapach, Strauss and Zhou (2010), even small values of R_{os}^2 may increase the utility of a CRRA investor by a considerable amount. Based on these concerns, the utility-based evaluation method which is proposed by Edison and Cho (1993), has been used widely in literature.

Assume in a market where the investment opportunity consists of a risk-free bond and a risky asset. A risk-averse investor constructs her portfolio by assigning different allocations between them. Since the only source of risk comes from the risky security, we can solve the allocation problem in a standard way. In the mean-

variance framework, the optimal investment strategy assigns the following weight (ω_t) to the risky asset

$$\omega_t = \frac{E_t(r_{t+1})}{\gamma \text{Var}_t(r_{t+1})},$$

where γ is the investor's Relative Risk Aversion coefficient, $E_t(r_{t+1})$ and $\text{Var}_t(r_{t+1})$ are the conditional expectation and conditional variance at time t , of r_{t+1} , the equity risk premium at time $t + 1$, respectively. The conditional variance is estimated by the historical variance, based on a ten-year rolling window. The conditional mean is forecasted by the model the investor chooses. Different models give different invest strategies, thus induce different utilities. It is natural to consider a model to be better if it results in higher utilities to investors. The idea of the economic evaluation is based on this fact, and compares the utility difference generated by two different models.

The resulting portfolio gross return, $R_{t+1,o}$, obtained by the optimal allocation is calculated as

$$R_{t+1,o} = \omega_t r_{t+1} + R_{t,f},$$

where r_{t+1} is the realized excess return of the risky security at time $t + 1$ and $R_{t,f} = 1 + r_{t,f}$ is the gross return of the risk-free asset from period t to $t + 1$.

Given the initial wealth W_0 and quadratic utility, the average utility over the out-of-sample period is given by

$$\bar{U} = \frac{W_0}{k} \sum_{t=0}^{k-1} (R_{t+1,o} - \frac{\gamma}{2(1+\gamma)} R_{t+1,o}^2). \quad (2.13)$$

For different models, we obtain different strategies, thus different average utilities. We prefer model with high average utility. The utility gain, comparing model A

with a benchmark, is given by

$$\Delta\bar{U} = 400\left(\frac{\bar{U}^A - \bar{U}^{bm}}{\bar{U}^{bm}}\right),$$

where \bar{U}^A is the average utility over the out-of-sample period generated by model A, and \bar{U}^{bm} is the respective value generated by the benchmark model A is compared with. The scalar 400 here is to express the utility gain to average annualized percentage return.

Following Edison and Cho (1993)'s idea, we also consider the wealth change in the mean-variance framework. Under the above set-up, consider an investor who has one unit initial wealth. At each time period she constructs her portfolio in the above way, then reinvests all her returns to the new portfolio she constructs in the next time period. In this way, her total wealth at time $m + k$ would be

$$W_T = \prod_{t=1}^k R_{m+t,A}, \quad (2.14)$$

where $R_{m+t,A}$ is the realized gross return she obtains if she employ model A. The wealth difference can be calculated by

$$\Delta W = \prod_{t=1}^k R_{m+t,A} - \prod_{t=1}^k R_{m+t,bm}. \quad (2.15)$$

Chapter 3

Theoretical Results

In this chapter, we present the theoretical results related with our model.

3.1 Autocorrelation Structure

In finance field, especially related with asset pricing, the data are mostly time series data. Therefore we need to first model the autocorrelation structure. We first consider the data generating process of the factors. Following the literature, we model the predictor x_t as a near unit root process, since most of the predictors are highly persistent. Specifically, following Cavanagh et al. (1995), we assume that the predictor x_t is a finite order autoregressive process:

$$x_t = \alpha_0 + v_t, \tag{3.1}$$

$$(1 - \alpha_1 L)b(L)v_t = \epsilon_{1t}, \tag{3.2}$$

where $b(L) = \sum_{i=0}^k b_i L^i$ with $b_0 = 1$. The roots of $b(L)$ are fixed and inside the unit circle. The above equation can be rewritten as:

$$\Delta x_t = \beta_0 + \beta_1 x_{t-1} + \zeta(L) \Delta x_{t-1} + \epsilon_{1t}, \quad (3.3)$$

where $\beta_0 = (1 - \alpha_1)b(1)\alpha_0$, $\beta_1 = (\alpha_1 - 1)b(1)$, and $\zeta(L) = -\sum_{j=1}^k L^{-1}[b_j - (1 - \alpha_1)\sum_{i=j}^k b_i]L^j$. By this representation, it is straightforward that x_t follows an $AR(k+1)$ process with the largest root $\frac{1}{\alpha_1}$. Following Wang's (2012) set-up, we consider the local-to-unity specification, in particular, $\alpha_1 = 1 + \frac{c}{T}$, where $c \leq 0$ and T is the sample size. When $c < 0$ is the mean reverting case while $c = 0$ generalizes a unit root process. The predictors such as earnings and economic ratios are modeled by this framework in the literature, due to their highly persistent while stationary properties.

By the functional central limit theory, similarly with Wang(2012), we can obtain the limit distribution of ϵ_{1t} .

Proposition 1. Using the above notation, assume that ϵ_{1t} is a martingale difference sequence with variance σ_1 , we have that

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} b^{-1}(L)\epsilon_{1t} \xrightarrow{d} Z_1(r) = \omega B(r), \quad (3.4)$$

where $\omega = \frac{\sigma_1^{\frac{1}{2}}}{b(1)}$, $B(r)$ is a Brownian motion, and $[\cdot]$ denotes the floor function.

Furthermore, the asymptotic properties of x_t can be generalized in a similar way. Define the diffusion process J_c by $dJ_c(s) = cJ_c(s)ds + dB_1(s)$ with initial condition $J_c(0) = 0$. We represent the demeaned variables by superscript μ . That is to say, $x^\mu = x_t - (T-1)^{-1}\sum_{t=2}^T x_{t-1}$ and $J_c^\mu(s) = J_c(s) - \int_0^s J_c(r)dr$. Following Phillips (1987), we have proposition 2.

Proposition 2. Using the above notation, we have

$$T^{-\frac{1}{2}}x_{[Tr]}^\mu \xrightarrow{d} \omega J_c^\mu, \quad (3.5)$$

$$T^{-2} \sum (x_{t-1}^\mu)^2 \xrightarrow{d} \int (\omega J_c^\mu)^2 dr, \quad (3.6)$$

$$T^{-1} \sum x_{t-1}^\mu b^{-1}(L)\epsilon_{1t} \xrightarrow{d} \int \omega J_c^\mu dZ_1, \quad (3.7)$$

when $T \rightarrow \infty$, where Z_1 denotes Brownian motion at time 1, which is the standard normal distribution.

The following standard result gives the asymptotic distribution of the t-statistic associated with the largest root of the autoregressive model for x_t in (3.16), when this root is modeled local-to-unity.

Proposition 3. The asymptotic representation of the standard t-statistics used to test $H_0 : \beta_1 = 0$ is given by

$$t_{\beta_1} \xrightarrow{d} \frac{c}{b(1)} \left[\int (J_c^\mu)^2 dr \right]^{\frac{1}{2}} + \frac{\int J_c^\mu dB_1}{\left[\int (J_c^\mu)^2 dr \right]^{\frac{1}{2}}}. \quad (3.8)$$

3.2 Quantile Regression Models

Next, we consider the relationship between x and y . Denote \mathcal{F}_t as the σ -algebra (information set) up to time t . The standard mean prediction model is given by

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + \epsilon_{2t}, \quad (3.9)$$

where $E[\epsilon_{2t} | \mathcal{F}_{t-1}] = 0$. In our framework, y_t is typically a financial return and x_t is a predictor whose predictive power is of interest. Under this classic set-up, the mean prediction is given by $E[y_t | \mathcal{F}_{t-1}] = \gamma_0 + \gamma_1 x_{t-1}$.

Let $F(\cdot)$ and $F_{t-1}(\cdot)$, which equals $P(\epsilon_{2t} < \cdot | \mathcal{F}_{t-1})$, denote the cumulative and conditional cumulative distribution functions of ϵ_{2t} , respectively. Define the τ th unconditional and conditional quantiles of ϵ_{2t} by $Q_{\epsilon_{2t}}(\tau) = F^{-1}(\tau)$ and $Q_{\epsilon_{2t}|\mathcal{F}_{t-1}}(\tau) = F_{t-1}^{-1}(\tau)$, respectively. In a quantile regression model, we allow x_{t-1} to impact not only the mean of y_t , but also the distribution of its error term, therefore, in addition to the following equation which holds in OLS set-ups,

$$Q_{y_t|\mathcal{F}_{t-1}}(\tau) = \gamma_0 + Q_{\epsilon_{2t}}(\tau) + \gamma_1 x_{t-1}, \quad (3.10)$$

we also have

$$Q_{\epsilon_{2t}|\mathcal{F}_{t-1}}(\tau) = \gamma_0(\tau) - \gamma_0 + (\gamma_1(\tau) - \gamma_1)x_{t-1}. \quad (3.11)$$

Combining the above two equations, we obtain the quantile regression model

$$Q_{y_t|\mathcal{F}_{t-1}}(\tau) = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1}. \quad (3.12)$$

The quantile regression model provides a flexible specification allowing the effect of x_{t-1} to be heterogeneous across the quantiles of y_t . We put the following assumptions of the error term in our framework.

Assumption 1. The error term ϵ_{2t} is a martingale difference sequence with variance σ_2 .

Assumption 2. ϵ_{1t} and ϵ_{2t} are independent.

Assumption 3. ϵ_{2t} has a continuous density function $f(\cdot)$, which is positive on $\{\epsilon_{2t} : 0 < F(\epsilon_{2t}) < 1\}$.

Assumption 4. The conditional distribution function $F_{t-1}(\cdot) = P(\epsilon_{2t} < \cdot | \mathcal{F}_{t-1})$ has derivative $f_{t-1}(\cdot)$ a.s.; $f_{t-1}(\cdot)$ is uniformly integrable for any sequence $s_n \rightarrow F^{-1}(\tau)$, and $E[f_{t-1}^\xi(F^{-1}(\tau))] < \infty$ for some $\xi > 1$.

The standard quantile regression coefficient estimates are given by

$$(\hat{\gamma}_0(\tau), \hat{\gamma}_1(\tau)) = \arg \min_{(\gamma_0, \gamma_1) \in \mathbb{R}^2} \sum_{t=1}^T \rho_\tau(y_t - \gamma_0 - \gamma_1 x_{t-1}), \quad (3.13)$$

where ρ is defined by $\rho_\tau(u) = u\psi_\tau(u)$ with $\psi_\tau(u) = \tau - I(u < 0)$.

Define $\epsilon_{2t\tau} = \epsilon_{2t} - F_{t-1}^{-1}(\tau) = y_t - \gamma_0(\tau) - \gamma_1(\tau)x_{t-1}$ and $Q_{\epsilon_{2t\tau}|\mathcal{F}_{t-1}}(\tau)$ as the τ th quantile of ϵ_{2t} conditional on \mathcal{F}_{t-1} , we may rewrite (3.9) as

$$y_t = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1} + \epsilon_{2t\tau} = \gamma(\tau)'z_{t-1} + \epsilon_{2t\tau}, \quad (3.14)$$

where $Q_{\epsilon_{2t\tau}|\mathcal{F}_{t-1}}(\tau) = 0$. By the definition of ψ , we have that the conditional expectation $E[\psi_\tau(\epsilon_{2t\tau})|\mathcal{F}_{t-1}]$ and the variance of the indicator function $I(\cdot)$ are $\tau(1 - \tau)$. The following proposition, comparable to Assumption A of Xiao (2009) is useful when we derive the asymptotic properties of parameter estimates.

Proposition 4. Under the above set-up, we have

$$T^{-\frac{1}{2}} \sum_{t=1}^{[T\tau]} \psi_\tau(\epsilon_{2t}) \xrightarrow{d} Z_\psi(r) = \sqrt{\tau(1 - \tau)}B(r), \quad (3.15)$$

where $B(r)$ is a Brownian motion. It follows that $T^{-1} \sum x_{t-1}^\mu \psi_\tau(\epsilon_{2t}) \xrightarrow{d} \int \omega J_c^\mu dZ_\psi$. The following theorem provides the limiting distribution of the predictive quantile regression estimator.

Theorem 1.(Wang, 2013) The asymptotic distribution of the regression coefficient estimates is

$$D_T(\hat{\gamma}(\tau) - \gamma(\tau)) \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))} \left[\int \bar{J}_c \bar{J}_c' \right]^{-1} \left[\int \bar{J}_c dZ_\psi \right]. \quad (3.16)$$

For the slope coefficient, in particular,

$$T(\hat{\gamma}_1(\tau) - \gamma_1(\tau)) \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))} \left[\int (\omega J_c^\mu)^2 \right]^{-1} \left[\int \omega J_c^\mu dZ_\psi \right], \quad (3.17)$$

where $D_T = \begin{pmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T \end{pmatrix}$ and $\bar{J}_c = (1, \omega J_c)'$.

The proof of Theorem 1 is given in Appendix B in Wang's paper (2012). The following corollary is a direct application of Theorem 1.

Corollary 1. The estimate of the τ th quantile of the distribution of $Y_t|X_{t-1}$,

$$\hat{Q}_{Y_t|X_{t-1}}(\tau|\mathcal{F}_{t-1}) = X_{t-1}^T \hat{\gamma}(\tau) \quad (3.18)$$

is consistent.

The ultimate forecast given by our method is a combination of different quantile predictions. Therefore, moving forward, we need to discuss the relationship between estimates with respect to different quantiles. Let $\{\tau_1, \tau_2, \dots, \tau_L\}$ be a set of quantiles with $\tau_1 \leq \tau_2 \leq \dots \leq \tau_L$. Accordingly, we introduce $\epsilon_{2t\tau_i}, \gamma(\tau_i), \psi_{\tau_i}$ and $Z_{\psi_{\tau_i}}$ as the variables and parameters with respect to different quantiles. The following two lemma will be needed for the theorem providing the asymptotic covariance.

Lemma 1.

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \begin{pmatrix} \psi_{\tau_i}(\epsilon_{2t}) \\ \psi_{\tau_j}(\epsilon_{2t}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_{\psi_{\tau_i}}(r) \\ Z_{\psi_{\tau_j}}(r) \end{pmatrix} = BM(0, \Omega_{ij}), \quad (3.19)$$

where $i < j$ and $\Omega_{ij} = \begin{pmatrix} \tau_i(1 - \tau_i) & \tau_i(1 - \tau_j) \\ \tau_i(1 - \tau_j) & \tau_j(1 - \tau_j) \end{pmatrix}$.

Proof. By functional central limit theorem,

$$\text{cov}(Z_{\psi_{\tau_i}}(r), Z_{\psi_{\tau_j}}(r)) = \text{cov}(\psi_{\tau_i}(\epsilon_{2t}), \psi_{\tau_j}(\epsilon_{2t})), \quad (3.20)$$

while since $\psi_{\tau_l} = \tau_l - I(\epsilon_{2t} < F^{-1}(\tau_l))$ for $l = i, j$, we have $E(\psi_{\tau_l}) = 0$. Therefore,

$$\begin{aligned} \text{cov}(\psi_{\tau_i}(\epsilon_{2t}), \psi_{\tau_j}(\epsilon_{2t})) &= E(\psi_{\tau_i}(\epsilon_{2t})\psi_{\tau_j}(\epsilon_{2t})) - E(\psi_{\tau_i}(\epsilon_{2t}))E(\psi_{\tau_j}(\epsilon_{2t})) \\ &= P(\epsilon_{2t} < F^{-1}(\tau_i))(\tau_i - 1)(\tau_j - 1) + P(F^{-1}(\tau_i) \leq \epsilon_{2t} < F^{-1}(\tau_j))\tau_i(\tau_j - 1) \\ &\quad + P(\epsilon_{2t} \geq F^{-1}(\tau_j))\tau_i\tau_j \\ &= \tau_i(\tau_i - 1)(\tau_j - 1) + (\tau_j - \tau_i)\tau_i(\tau_j - 1) + (1 - \tau_j)\tau_i\tau_j \\ &= \tau_i(1 - \tau_j). \end{aligned}$$

By Proposition 4, we have that $T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor Tr \rfloor} \psi_{\tau_i}(\epsilon_{2t\tau_i}) \rightarrow Z_{\psi_{\tau_i}}(r) = \sqrt{\tau_i(1 - \tau_i)}B(r)$ and $T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor Tr \rfloor} \psi_{\tau_j}(\epsilon_{2t\tau_j}) \rightarrow Z_{\psi_{\tau_j}}(r) = \sqrt{\tau_j(1 - \tau_j)}B(r)$. Combining the results we get the conclusion. \square

Lemma 2. Given that X is independent of both Y_1 and Y_2 , and $E(Y_1) = E(Y_2) = 0$, we have that

$$\text{cov}(XY_1, XY_2) = EX^2 \text{cov}(Y_1, Y_2). \quad (3.21)$$

Proof.

$$\begin{aligned} \text{cov}(XY_1, XY_2) &= E(XY_1XY_2) - E(XY_1)E(XY_2) \\ &= EX^2E(Y_1Y_2) - (EX)^2E(Y_1)E(Y_2) \\ &= EX^2E(Y_1Y_2) \\ &= EX^2 \text{cov}(Y_1, Y_2) \end{aligned}$$

□

Then we have the following theorem.

Theorem 2. The joint asymptotic distribution of the regression coefficient estimates is

$$MD_T \begin{pmatrix} \hat{\gamma}(\tau_1) - \gamma(\tau_1) \\ \hat{\gamma}(\tau_2) - \gamma(\tau_2) \\ \dots \\ \hat{\gamma}(\tau_L) - \gamma(\tau_L) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{f(F^{-1}(\tau_1))} [\int \bar{J}_c \bar{J}'_c]^{-1} [\int \bar{J}_c dZ_{\psi_{\tau_1}}] \\ \frac{1}{f(F^{-1}(\tau_2))} [\int \bar{J}_c \bar{J}'_c]^{-1} [\int \bar{J}_c dZ_{\psi_{\tau_2}}] \\ \dots \\ \frac{1}{f(F^{-1}(\tau_L))} [\int \bar{J}_c \bar{J}'_c]^{-1} [\int \bar{J}_c dZ_{\psi_{\tau_L}}] \end{pmatrix}, \quad (3.22)$$

where $MD_T = I_L \otimes D_T$. The asymptotic variance matrix is $\Omega = (\omega_{ij})_{L \times L}$ with

$$\omega_{ij} = \begin{cases} \frac{1}{f(F^{-1}(\tau_i))} \frac{1}{f(F^{-1}(\tau_j))} \tau_i (1 - \tau_j) \int E[(\int \bar{J}_c \bar{J}'_c)^{-1} \bar{J}_c(s)]^2 ds, & \text{when } i \leq j \\ \frac{1}{f(F^{-1}(\tau_i))} \frac{1}{f(F^{-1}(\tau_j))} \tau_j (1 - \tau_i) \int E[(\int \bar{J}_c \bar{J}'_c)^{-1} \bar{J}_c(s)]^2 ds, & \text{when } i > j \end{cases}.$$

Proof. Noting that

$$\int \bar{J}_c dZ_{\psi_{\tau_i}} = \lim_{\Delta s \rightarrow 0} \sum \bar{J}_c(s_n) (Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})). \quad (3.23)$$

For notation simplicity, denote $K = [\int \bar{J}_c \bar{J}'_c]^{-1}$. Consider the covariance between the summations $cov(K \sum \bar{J}_c(s_n) (Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})), K \sum \bar{J}_c(s_n) (Z_{\psi_{\tau_j}}(s_n) - Z_{\psi_{\tau_j}}(s_{n-1})))$.

By lemma 1, $\begin{pmatrix} Z_{\psi_{\tau_i}}(r) \\ Z_{\psi_{\tau_j}}(r) \end{pmatrix}$ is a two dimensional Brownian motion, therefore, we have that the expectation of the increments is zero as well as that the increments with respect to non-overlapping time periods are independent with each other. Thus we

have

$$\begin{aligned}
& cov(K\bar{J}_c(s_n)(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})), K\bar{J}_c(s_m)(Z_{\psi_{\tau_i}}(s_m) - Z_{\psi_{\tau_i}}(s_{m-1}))) \\
&= E(K^2(\bar{J}_c(s_n)(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1}))\bar{J}_c(s_m)(Z_{\psi_{\tau_i}}(s_m) - Z_{\psi_{\tau_i}}(s_{m-1})))) \\
&\quad - E(K\bar{J}_c(s_n)(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})))E(K\bar{J}_c(s_m)(Z_{\psi_{\tau_i}}(s_m) - Z_{\psi_{\tau_i}}(s_{m-1}))) \\
&= E(K^2\bar{J}_c(s_n)\bar{J}_c(s_m))E(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1}))E(Z_{\psi_{\tau_i}}(s_m) - Z_{\psi_{\tau_i}}(s_{m-1})) \\
&\quad - E(K\bar{J}_c(s_n))E(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1}))E(K\bar{J}_c(s_m))E(Z_{\psi_{\tau_i}}(s_m) - Z_{\psi_{\tau_i}}(s_{m-1}))) \\
&= 0,
\end{aligned}$$

for $m \neq n$.

Therefore we have

$$\begin{aligned}
& cov(K \sum \bar{J}_c(s_n)(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})), K \sum \bar{J}_c(s_n)(Z_{\psi_{\tau_j}}(s_n) - Z_{\psi_{\tau_j}}(s_{n-1}))) \\
&= \sum cov(K\bar{J}_c(s_n)(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})), K\bar{J}_c(s_n)(Z_{\psi_{\tau_j}}(s_n) - Z_{\psi_{\tau_j}}(s_{n-1}))) \\
&= \sum E(K\bar{J}_c(s_n))^2 cov(Z_{\psi_{\tau_i}}(s_n) - Z_{\psi_{\tau_i}}(s_{n-1})), (Z_{\psi_{\tau_j}}(s_n) - Z_{\psi_{\tau_j}}(s_{n-1}))) \\
&= \begin{cases} \sum E(K\bar{J}_c(s_n))^2 \tau_i(1 - \tau_j) \Delta s, & \text{when } i \leq j \\ \sum E(K\bar{J}_c(s_n))^2 \tau_j(1 - \tau_i) \Delta s, & \text{when } i > j \end{cases} \\
&\rightarrow \begin{cases} \int E(K\bar{J}_c(s))^2 \tau_i(1 - \tau_j) ds, & \text{when } i \leq j \\ \int E(K\bar{J}_c(s))^2 \tau_j(1 - \tau_i) ds, & \text{when } i > j \end{cases}.
\end{aligned} \tag{3.24}$$

The first equation is due to the above argument; the second one is based on Assumption 2 and lemma 2; the third equation is derived from the lemma 1 and the last one is by the definition of Riemann Integral. Reorganize the items and multiply the constant we get the covariance matrix. Combining with the results in Theorem

1, we prove the conclusion. □

3.3 Markov Chain

Moving forward, we consider the set up of state variable S_t . Let $\tau_1 < \tau_2 < \tau_3$ be three quantiles with given value. Separate the range of $Y_t|X_{t-1}$ into three parts: $(-\infty, Q_{Y_t|X_{t-1}}(\tau_1))$, $[Q_{Y_t|X_{t-1}}(\tau_1), Q_{Y_t|X_{t-1}}(\tau_3))$ and $[Q_{Y_t|X_{t-1}}(\tau_3), +\infty)$. S_t , taking values 1, 2, and 3, denotes the interval in which Y_t falls. We have the following assumptions for S_t .

Assumption 5. S_t is exogenous, which means, S_t is independent with X_t process.

Assumption 6. S_t follows a three state first order Markov Chain.

Assumption 7. S_t is irreducible and aperiodic, therefore ergodic.

Assumption 5 assures the exogeneity of S_t . This makes economic sense, since S_t denotes the state of the whole market, thus a macro-index. The predictors, such as economic ratios, are micro-indices changing independently with market states. Assumption 6 is consistent with literature (e.g., Goutte(2012), Bai and Wang (2010)). Assumption 7 imposes a mild restriction on the structure of the state variable. By Assumption 6, we see that the state space $E = \{1, 2, 3\}$, which is finite. For a finite state Markov chain, which is recurrent as well as irreducible and aperiodic tends to be ergodic. We know that the market moving trend is difficult to predict; the existence of such huge noise means that it is reasonable for us to assume that the transition probabilities of the Markov chain is non-zero. Therefore by our definition of the state variable, we can see that each state is communicated with the others, and $P_{ii} \neq 0, \forall i$. By the above argument, we can see that the ergodic property is a natural assumption for market state variables.

By Assumption 7, we conclude that the limiting distribution of S_t exists. Implied

by the definition, the limiting distribution of S_t is the following

$$S_t = \begin{cases} 1, & \text{with probability } \tau_1 \\ 2, & \text{with probability } \tau_3 - \tau_1 \\ 3, & \text{with probability } 1 - \tau_3 \end{cases} .$$

The estimation of the state variable S_t , which is given by

$$\hat{S}_t = \begin{cases} 1, & \text{when } y_t < \hat{Q}_{Y_t|X_{t-1}}(\tau_1) \\ 3, & \text{when } y_t \geq \hat{Q}_{Y_t|X_{t-1}}(\tau_3) \\ 2, & \text{otherwise} \end{cases}$$

can be rewritten as

$$\begin{aligned} \hat{S}_t &= 1I(y_t < \hat{Q}_{Y_t|X_{t-1}}(\tau_1)) + 3I(y_t \geq \hat{Q}_{Y_t|X_{t-1}}(\tau_3)) \\ &\quad + 2(1 - I(y_t < \hat{Q}_{Y_t|X_{t-1}}(\tau_1)) - I(y_t \geq \hat{Q}_{Y_t|X_{t-1}}(\tau_3))). \end{aligned}$$

Denote $\hat{I}_1 = I(y_t < \hat{Q}_{Y_t|X_{t-1}}(\tau_1))$ and $\hat{I}_3 = I(y_t \geq \hat{Q}_{Y_t|X_{t-1}}(\tau_3))$, we have

$$\begin{aligned} \hat{S}_t &= \hat{I}_1 + 3\hat{I}_3 + 2(1 - \hat{I}_1 - \hat{I}_3) \\ &= 2 - \hat{I}_1 + \hat{I}_3. \end{aligned}$$

Let

$$S_t = \begin{cases} 1, & \text{when } y_t < Q_{Y_t|X_{t-1}}(\tau_1) \\ 3, & \text{when } y_t \geq Q_{Y_t|X_{t-1}}(\tau_3) \\ 2, & \text{otherwise} \end{cases}$$

The following theorem shows that \hat{S}_t converges to S_t in probability.

Theorem 3. Using the notations above, we have that

$$\lim_{T \rightarrow +\infty} E|\hat{S}_t - S_t| = 0, \quad (3.25)$$

where T denotes the sample length of the data used for parameter estimation.

Proof. Denote $I_1 = I(y_t < Q_{Y_t|X_{t-1}}(\tau_1))$ and $I_3 = I(y_t \geq Q_{Y_t|X_{t-1}}(\tau_3))$, by the same argument, we have

$$S_t = 2 - I_1 + I_3. \quad (3.26)$$

Let's first show that \hat{I}_1 is L_1 -convergent to I_1 .

$$\hat{I}_1 - I_1 = \begin{cases} 1, & \text{when } Q_{Y_t|X_{t-1}}(\tau_1) \leq y_t < \hat{Q}_{Y_t|X_{t-1}}(\tau_1) \\ -1, & \text{when } \hat{Q}_{Y_t|X_{t-1}}(\tau_1) \leq y_t < Q_{Y_t|X_{t-1}}(\tau_1) \\ 0, & \text{otherwise} \end{cases} .$$

Therefore

$$E|\hat{I}_1 - I_1| = P(y_t \in [\min\{Q_{Y_t|X_{t-1}}(\tau_1), \hat{Q}_{Y_t|X_{t-1}}(\tau_1)\}, \max\{Q_{Y_t|X_{t-1}}(\tau_1), \hat{Q}_{Y_t|X_{t-1}}(\tau_1)\}]).$$

Denote the interval as A .

From corollary 1, we have that $\hat{Q}_{Y_t|X_{t-1}}(\tau_1)$ is a consistent estimator of $Q_{Y_t|X_{t-1}}(\tau_1)$, which means that, for any $\epsilon > 0$ and $\delta > 0$, there exists N , such that for any $T > N$, $P(|\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1) - Q_{Y_t|X_{t-1}}(\tau_1)| > \epsilon) < \frac{\delta}{2}$, where $\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1)$ is the estimator of $Q_{Y_t|X_{t-1}}(\tau_1)$ based on T time periods. Let's denote the interval $[Q_{Y_t|X_{t-1}}(\tau_1) -$

$\epsilon, Q_{Y_t|X_{t-1}}(\tau_1) + \epsilon]$ as B and its complement as C . Therefore,

$$\begin{aligned} P(y_t \in A) &= P(\{y_t \in A\} \cap \{\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1) \in C\}) + P(\{y_t \in A\} \cap \{\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1) \in B\}) \\ &\leq P(\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1) \in C) + P(\{y_t \in A\} \cap \{\hat{Q}_{Y_t|X_{t-1}}^{(T)}(\tau_1) \in B\}) \\ &\leq \frac{\delta}{2} + \max\{F(Q_{Y_t|X_{t-1}} + \epsilon) - F(Q_{Y_t|X_{t-1}}), F(Q_{Y_t|X_{t-1}}) - F(Q_{Y_t|X_{t-1}} - \epsilon)\}. \end{aligned}$$

δ can be set as close to zero as needed, and by assumption 3 and 4, the second item decreases to zero as ϵ tends to zero. Therefore, we can set ϵ small enough such that

$$\max\{F(Q_{Y_t|X_{t-1}} + \epsilon) - F(Q_{Y_t|X_{t-1}}), F(Q_{Y_t|X_{t-1}}) - F(Q_{Y_t|X_{t-1}} - \epsilon)\} \leq \frac{\delta}{2}. \quad (3.27)$$

So we have shown that

$$\lim_{T \rightarrow +\infty} E|\hat{I}_1 - I_1| = 0. \quad (3.28)$$

Similarly, we have

$$\lim_{T \rightarrow +\infty} E|\hat{I}_3 - I_3| = 0. \quad (3.29)$$

So

$$\begin{aligned} E|\hat{S}_t - \hat{S}| &= E|(2 - \hat{I}_2 + \hat{I}_3) - (2 - I_1 + I_3)| \\ &= E|\hat{I}_3 - I_3 - (\hat{I}_1 - I_1)| \\ &\leq E|\hat{I}_1 - I_1| + E|\hat{I}_3 - I_3| \end{aligned}$$

also converges to zero. □

Next we consider the transition probabilities for S_t : $P_{ij} = P(S_{t+1} = j | S_t = i)$.

The estimation is given by:

$$\hat{P}_{ij} = \frac{\sum_{t=1}^{T-1} I(\hat{S}_{t+1} = j \text{ and } \hat{S}_t = i)}{\sum_{t=1}^{T-1} I(\hat{S}_t = i)}. \quad (3.30)$$

The following theorem shows the consistency of \hat{P}_{ij} .

Theorem 4. Given that \hat{S}_t is a consistent estimator of S_t , we have

$$\hat{P}_{ij} \xrightarrow{p} P_{ij}, \quad (3.31)$$

where \hat{P}_{ij} is calculated by equation (3.30).

Proof. Denote π_i as the limiting probability of state i . By definition, we have

$$\frac{\sum I(S_t = i)}{T-1} \xrightarrow{p} \pi_i. \quad (3.32)$$

Also, we have

$$\begin{aligned} E \left| \frac{\sum I(S_t = i)}{T-1} - \frac{\sum I(\hat{S}_t = i)}{T-1} \right| &\leq \frac{1}{T-1} \sum E |I(S_t = i) - I(\hat{S}_t = i)| \\ &\leq \frac{1}{T-1} \sum P(S_t \neq \hat{S}_t) \rightarrow 0, \end{aligned}$$

by the probability convergence of \hat{S}_t . Therefore, we have that

$$\frac{\sum I(\hat{S}_t = i)}{T-1} \xrightarrow{p} \pi_i. \quad (3.33)$$

Now consider

$$\frac{\sum I(S_{t+1} = j \text{ and } S_t = i)}{T-1} \xrightarrow{p} P(S_{t+1} = j \text{ and } S_t = i) = P_{ij}\pi_i.$$

This can be obtained by considering $\{S_t, S_{t+1}\}$ as a nine state Markov chain with limiting probabilities $P_{ij}\pi_i$. Similarly, we have that

$$\begin{aligned} & E \left| \frac{\sum I(S_{t+1} = j \text{ and } S_t = i)}{T-1} - \frac{\sum I(\hat{S}_{t+1} = j \text{ and } S_t = i)}{T-1} \right| \\ & \leq \frac{1}{T-1} \sum E |I(S_{t+1} = j \text{ and } S_t = i) - I(\hat{S}_{t+1} = j \text{ and } S_t = i)| \\ & \leq \frac{1}{T-1} \sum (P(S_t \neq \hat{S}_t) + P(S_{t+1} \neq \hat{S}_{t+1})) \rightarrow 0, \end{aligned}$$

also by the probability convergence of \hat{S}_t . Thus we have

$$\frac{\sum I(\hat{S}_{t+1} = j \text{ and } \hat{S}_t = i)}{T-1} \xrightarrow{p} P_{ij}\pi_i. \quad (3.34)$$

By Slutsky's Theorem, we have that

$$\hat{P}_{ij} \xrightarrow{p} \frac{P_{ij}\pi_i}{\pi_i} = P_{ij}. \quad (3.35)$$

□

To explore further the property of the estimation of transition probability, we first extend S_t into a nine state Markov chain, $\{(S_t, S_{t+1})\} = \{X_t\}$. The following lemma shows the property of this new Markov chain.

Lemma 3. $\{X_t\}$ is also ergodic, and hence has stationary distribution, which is given by

$$\pi(i, j) = P_{ij}\pi_i, \quad (3.36)$$

where P_{ij} and π_i are the transition probabilities and the limiting probabilities of S_t , respectively. The limiting distribution is defined in the state space of X_t :

$$E' = \{(i, j), i, j = 1, 2, 3\}.$$

Proof. The ergodic property is easy to show after writing down the transition probability matrix of X_t in terms of P_{ij} , which is

$$\begin{pmatrix} & (1, 1) & (1, 2) & (1, 3) & (2, 1) & (2, 2) & (2, 3) & (3, 1) & (3, 2) & (3, 3) \\ (1, 1) & P_{11} & P_{12} & P_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ (1, 2) & 0 & 0 & 0 & P_{21} & P_{22} & P_{23} & 0 & 0 & 0 \\ (1, 3) & 0 & 0 & 0 & 0 & 0 & 0 & P_{31} & P_{32} & P_{33} \\ (2, 1) & P_{11} & P_{12} & P_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ (2, 2) & 0 & 0 & 0 & P_{21} & P_{22} & P_{23} & 0 & 0 & 0 \\ (2, 3) & 0 & 0 & 0 & 0 & 0 & 0 & P_{31} & P_{32} & P_{33} \\ (3, 1) & P_{11} & P_{12} & P_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ (3, 2) & 0 & 0 & 0 & P_{21} & P_{22} & P_{23} & 0 & 0 & 0 \\ (3, 3) & 0 & 0 & 0 & 0 & 0 & 0 & P_{31} & P_{32} & P_{33} \end{pmatrix}.$$

From the matrix, we can see that $\forall (i, j), (i', j') \in E'$,

$$(i, j) \rightleftharpoons (i', j') \iff \begin{cases} i \rightleftharpoons j \\ i' \rightleftharpoons j' \\ j \rightleftharpoons i' \end{cases}. \quad (3.37)$$

Hence we can see that the irreducible property of X_t is promised by that of S_t .

What is more,

$$P_{jj} \neq 0 \implies P_{(j,j),(j,j)} \neq 0.$$

Therefore the aperiodic property can also be deduced by that of S_t . Furthermore,

since X_t is a finite state Markov chain, we can see that it is ergodic.

The expression of the limiting distribution can be directly obtained by its definition and the limiting distribution of S_t . \square

To show the normality of the transition probability estimations, in addition to consistence, we need further acquire the rate of the convergence. So here we first state one more assumption to control the rate of the convergence of S_t .

Assumption 8. $\{S_t\}$ is geometrically ergodic, which is to say,

$$\forall \text{ initial distribution } \lambda_0 \text{ on } E, \|P^t(\lambda_0, \cdot) - \pi(\cdot)\| \leq M\rho^t,$$

where M is a finite positive number and $0 < \rho < 1$. Assumption 8 seems to put further restrictions on S_t , however, since S_t is a finite state Markov chain, ergodic is equivalent with geometrically ergodic. Therefore, Assumption 8 contains no further information than Assumption 7. Stating it here individually is aimed to explicitly show the property we base on at each step.

The following theorem holds once we have assumption 8.

Theorem 5. Let S_t be a geometrically ergodic Markov chain, we have that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T (I(S_t = i, S_{t+1} = j) - \pi_i P_{ij}) \xrightarrow{d} N(0, \sigma_f^2), \quad (3.38)$$

where

$$\sigma_f^2 = \text{Var}(f(X_0)) + 2 \sum_{t=1}^{\infty} \text{Cov}(f(X_0), f(X_t)) \quad (3.39)$$

and

$$f(X_t) = I_{(i,j)}(X_t). \quad (3.40)$$

Remark. This theorem is a direct application of GL.Jones's work in 2004, Theorem 2 and Theorem 5, since $f(X_t)$ is a Borel function of X_t . The limiting variance is

finite since,

$$\begin{aligned}
& \sum_{t=1}^{\infty} Cov(f(X_0), f(X_t)) \\
&= \sum_{t=1}^{\infty} (E(I_{(i,j)}(X_0) \cdot I_{(i,j)}(X_t)) - EI_{(i,j)}(X_0)EI_{(i,j)}(X_t)) \\
&= \sum_{t=1}^{\infty} (P[X_0 = (i,j), X_t = (i,j)] - P[X_0 = (i,j)]P[X_t = (i,j)]) \\
&= \sum_{t=1}^{\infty} P[X_0 = (i,j)](P[X_t = (i,j)|X_0 = (i,j)] - P[X_t = (i,j)]).
\end{aligned}$$

For t large enough, $P[X_t = (i,j)] \approx \pi(i,j)$, therefore by geometric ergodic property, we have

$$\sum_{t=1}^{\infty} Cov(f(X_0), f(X_t)) \leq \sum_{t=1}^{\infty} P[X_0 = (i,j)] \cdot M \cdot \rho^t < \infty.$$

By theorem 5, we can see that if we estimate the transition probability by the true empirical estimation, given by

$$\tilde{P}_{ij} = \frac{\sum_{t=1}^{T-1} I(S_{t+1} = j \text{ and } S_t = i)}{\sum_{t=1}^{T-1} I(S_t = i)}, \quad (3.41)$$

we will get asymptotic normality. Notice that this expression is obtained by replacing \hat{S}_t with S_t in the estimation of P_{ij} given by (3.43). The following lemma provides us with a bridge to connect the normality of \tilde{P}_{ij} with \hat{P}_{ij} .

Lemma 4. Let $\{X_t\}, \{Y_t\}$ be two sequences of random variables, assume

$$\begin{aligned}
X_t &\xrightarrow{d} X_0, \\
|Y_t - X_t| &\xrightarrow{p} 0,
\end{aligned}$$

then we have

$$Y_t \xrightarrow{d} X_0. \quad (3.42)$$

Proof. To show that $Y_t \xrightarrow{d} X_0$, equivalently, we can show that $\forall x \in C(F_{X_0})$, $F_{Y_t}(x) \rightarrow F_{X_0}(x)$, i.e., $P(Y_t \leq x) \rightarrow P(X_0 \leq x)$, where F are the cumulative distribution functions and $C(\cdot)$ denotes the continuous points of a function.

Let $E_t = \{|Y_t - X_t| > \epsilon\}$, and $\forall x \in C(F_{X_0})$, we have

$$\begin{aligned} P(Y_t \leq x) &= P(Y_t \leq x; E_t) + P(Y_t \leq x; E_t^c) \\ &\leq P(E_t) + P((Y_t \leq x; E_t^c), \end{aligned}$$

when $(Y_t, X_t) \in E_t^c = \{|Y_t - X_t| \leq \epsilon\}$, $Y_t \leq x$ implies $X_t \leq x + \epsilon$. Therefore, we have that

$$P(Y_t \leq x) \leq P(E_t) + P(X_t \leq x + \epsilon).$$

Taking *limsup* on both hand sides, we obtain

$$\limsup_{t \rightarrow \infty} P(Y_t \leq x) \leq 0 + P(X_0 \leq x + \epsilon).$$

Let $\epsilon \searrow 0$, by continuity of F_{X_0} at x ,

$$\limsup_{t \rightarrow \infty} P(Y_t \leq x) \leq P(X_0 \leq x). \quad (3.43)$$

On the other hand, we have

$$P(Y_t \leq x) \geq P(Y_t \leq x; E_t^c),$$

and when $(Y_t, X_t) \in E_t^c$, $X_t \leq x - \epsilon$ implies $Y_t \leq x$, therefore, we have

$$\begin{aligned} P(Y_t \leq x) &\geq P(X_t \leq x - \epsilon; E_t^c) \\ &\geq P(X_t \leq x - \epsilon) - P(E_t). \end{aligned}$$

The second inequality is because

$$\begin{aligned} &P(X_t \leq x - \epsilon; E_t^c) \\ &= P(X_t \leq x - \epsilon) - P(X_t \leq x - \epsilon \cap E_t) \\ &\geq P(X_t \leq x - \epsilon) - P(E_t). \end{aligned}$$

Take *liminf* on both hand sides, we have that

$$\liminf_{t \rightarrow \infty} P(Y_t \leq x) \geq P(X_0 \leq x - \epsilon) - 0,$$

and let $\epsilon \searrow 0$, again, by continuity of F_{X_0} at x , we obtain

$$\liminf_{t \rightarrow \infty} P(Y_t \leq x) \geq P(X_0 \leq x). \quad (3.44)$$

Combining (3.43) and (3.44) we obtain that $\forall x \in C(F_{X_0})$,

$$\lim_{t \rightarrow \infty} P(Y_t \leq x) = P(X_0 \leq x). \quad (3.45)$$

Thus we complete the proof of the lemma. \square

With the conclusion of Lemma 4, now we are only one step to the normality property of transition probability estimation. The following Lemma 5 fills this gap.

Lemma 5. In the above set up, we have

$$\frac{1}{\sqrt{T}} \left(\sum_{t=1}^T I(\hat{S}_t = i, \hat{S}_{t+1} = j) - \sum_{t=1}^T I(S_t = i, S_{t+1} = j) \right) \xrightarrow{p} 0. \quad (3.46)$$

Proof. By (3.16) in Theorem 1, we have that the convergence rate for the quantile regression coefficients $\hat{\gamma}(\tau)$ is

$$\begin{pmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T \end{pmatrix}.$$

Since $\hat{Q}_{Y_t|X_{t-1}}(\tau|\mathcal{F}_{t-1}) = X_{t-1} \cdot \hat{\gamma}(\tau)$, we can obtain a bound for the convergence rate of $\hat{Q}_{Y_t|X_{t-1}}(\tau|\mathcal{F}_{t-1})$, which is $T^{\frac{1}{2}}$. This leads us to the result that, $\forall \epsilon > 0$,

$$P(|\hat{Q}_{Y_t|X_{t-1}}(\tau|\mathcal{F}_{t-1}) - Q_{Y_t|X_{t-1}}(\tau|\mathcal{F}_{t-1})| > \epsilon) \leq \frac{1}{T} \cdot M_\epsilon, \quad (3.47)$$

the above inequality is a straightforward application of Chebyshev's Inequality, $P(|X - EX| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$.

Recall that the estimations and the true values of state variable S_t is given by

$$\begin{aligned} \hat{S}_t &= 1I(Y_t < \hat{Q}(\tau_1)) + 2I(Y_t \in [\hat{Q}(\tau_1), \hat{Q}(\tau_3)]) + 3I(Y_t \geq \hat{Q}(\tau_3)), \\ S_t &= 1I(Y_t < Q(\tau_1)) + 2I(Y_t \in [Q(\tau_1), Q(\tau_3)]) + 3I(Y_t \geq Q(\tau_3)), \end{aligned}$$

where for notation simplicity, $\hat{Q}(\tau_i)$ and $Q(\tau_i)$ are abbreviations for $\hat{Q}_{Y_t|X_{t-1}}(\tau_i|\mathcal{F}_{t-1})$

and $Q_{Y_t|X_{t-1}}(\tau_i|\mathcal{F}_{t-1})$ respectively. Thus we have

$$\begin{aligned}
P(\hat{S}_t \neq S_t) &= P(|\hat{Q}(\tau_1) - Q(\tau_1)| > \epsilon \cup |\hat{Q}(\tau_3) - Q(\tau_3)| > \epsilon \text{ for some } \epsilon > 0) \\
&\leq P(|\hat{Q}(\tau_1) - Q(\tau_1)| > \epsilon \text{ for some } \epsilon > 0) + P(|\hat{Q}(\tau_3) - Q(\tau_3)| > \epsilon \text{ for some } \epsilon > 0) \\
&\leq 2 \cdot \frac{1}{T} \cdot M_\epsilon.
\end{aligned}$$

So we get that

$$\begin{aligned}
&E \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T I(\hat{S}_t = i, \hat{S}_{t+1} = j) - \frac{1}{\sqrt{T}} \sum_{t=1}^T I(S_t = i, S_{t+1} = j) \right| \\
&\leq \frac{1}{\sqrt{T}} \sum_{t=1}^T P[(\hat{S}_t, \hat{S}_{t+1}) \neq (S_t, S_{t+1})] \\
&\leq \frac{1}{\sqrt{T}} \sum_{t=1}^T [P(\hat{S}_t \neq S_t) + P(\hat{S}_{t+1} \neq S_{t+1})] \\
&\leq \frac{1}{\sqrt{T}} \sum_{t=1}^T 2 \cdot 2 \cdot \frac{1}{T} \cdot M_\epsilon \\
&= \frac{1}{\sqrt{T}} \cdot 4M_\epsilon \rightarrow 0 \quad \text{as } T \rightarrow \infty.
\end{aligned}$$

The above inequality holds for any $\epsilon > 0$. Therefore we showed the convergence in probability conclusion in the lemma. \square

Combining the results in Theorem 5, Lemma 4 and Lemma 5, we immediately get the following

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T (I(\hat{S}_t = i, \hat{S}_{t+1} = j) - \pi_i P_{ij}) \xrightarrow{d} N(0, \sigma_f^2), \quad (3.48)$$

where σ_f^2 is defined by (3.39) and (3.40).

By similar argument in the proof of Theorem 4, and the application of ergodic

theorem, we also have

$$\frac{1}{T} \sum_{t=1}^T I(\hat{S}_t = i) \xrightarrow{p} \pi_i \quad (3.49)$$

Combining (3.48) and (3.49) we obtain the normality property, which is given by Theorem 6.

Theorem 6. With all the assumptions and notations above, we have

$$\sqrt{T}(\hat{P}_{ij} - P_{ij}) = \sqrt{T} \left(\frac{\sum_{t=1}^{T-1} I(\hat{S}_{t+1} = j, \hat{S}_t = i)}{\sum_{t=1}^{T-1} I(\hat{S}_t = i)} - P_{ij} \right) \xrightarrow{d} N(0, \sigma^2) \quad (3.50)$$

where $\sigma^2 = \frac{\sigma_f^2}{\pi_i^2}$.

In this chapter, we represent the primary theoretical results related with our model. First, based on the property of economic data, we construct the autocorrelation structure of the dependent variables. An AR model is applied to capture the evolution dynamics of economic factors. Next, we show the consistency of the parameter estimation in a quantile regression model with time series data. Furthermore, the asymptotic covariance between the parameter estimation of different quantiles is given.

In addition, with some mild assumptions, we show that the estimates of the underlying Markov chain which controls the movement of the market state variable are consistent. Particularly, the estimations of the state variables and the transition probabilities given in chapter 2 are consistent, and we further obtain the asymptotic normality property.

Chapter 4

Numerical Study

Theorem 2 in the last chapter shows that the estimates of the coefficients in our quantile regression method is consistent. However, in order to further generate statistical inference, we need additional properties such as asymptotic normality of the estimates. Unfortunately, the form of the asymptotic distribution is irregular unless in some very rare cases. Therefore in this chapter, we present the simulation studies of the time-series data quantile regression to further explore the properties of the estimates.

Let's first come back to an OLS set up, where the linear relationship between the response variable y and independent variable x can be modeled by

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + \epsilon_t. \quad (4.1)$$

In this standard set up, we have the assumption that x is independent with the error term ϵ . By such independence, we can see that x imposes the same influence on different quantiles of y . That is to say, if equation (4.1) is our true model, which means, we generate the response variable y according to this relationship, then the

true value of the slope remains to be γ_1 for all of the quantiles while the true value of the intercept shifts from γ_0 by the amount that determined both by the quantile we estimated on and the distribution of ϵ . Actually, given \mathcal{F}_t , the τ th quantile of y_t is

$$Q_{y_t|\mathcal{F}_{t-1}}(\tau) = \gamma_0 + \gamma_1 x_{t-1} + Q_{\epsilon_t}(\tau). \quad (4.2)$$

Therefore we obtain,

$$\gamma_0(\tau) = \gamma_0 + F^{-1}(\tau), \quad (4.3)$$

$$\gamma_1(\tau) = \gamma_1, \quad (4.4)$$

where F is the cumulative distribution function of the error term ϵ_t given the information up to time $t - 1$.

The reason why we introduce quantile regression as an alternative to OLS is that in some cases, the above independence assumption does not hold, that is to say, x can also influence the distribution of the error term. In particular, we model the conditional quantile of the error term as

$$Q_{\epsilon_t|\mathcal{F}_{t-1}}(\tau) = \gamma_0(\tau) - \gamma_0 + (\gamma_1(\tau) - \gamma_1)x_{t-1}, \quad (4.5)$$

then we will have

$$Q_{y_t|\mathcal{F}_{t-1}}(\tau) = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1}. \quad (4.6)$$

From this argument, we can see that in a quantile regression model, the independent variable x can influence the distribution of the error term, thus has different contribution to different quantiles of y .

In this simulation study, we assume that x has linear influence in the error term,

$$\epsilon_t = (\delta_0 + \delta_1 x_{t-1})u_t, \quad (4.7)$$

where u_t are *i.i.d.* errors, and $\delta_0 \neq 0$. Combining this assumption with (4.1), we obtain that

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + (\delta_0 + \delta_1 x_{t-1})u_t. \quad (4.8)$$

Under this case, the true value of the parameters in (4.6) is given by,

$$\gamma_0(\tau) = \gamma_0 + \delta_0 G^{-1}(\tau), \quad (4.9)$$

$$\gamma_1(\tau) = \gamma_1 + \delta_1 G^{-1}(\tau), \quad (4.10)$$

where G is the cumulative distribution function of u . Notice that if we set

$$\delta_1 = 0, \quad (4.11)$$

then the error term is independent of x_t , thus we obtain an OLS set up. Other than this circumstance, both $\gamma_0(\tau)$ and $\gamma_1(\tau)$ depend non-trivially on τ .

In the simulation study, the independent variable x is generated by (3.1) and (3.2). For convenience reason, we set $x_t = (1 + \frac{c}{T})x_{t-1} + \epsilon_{1t}$, where ϵ_{1t} are *i.i.d.* errors independent with u_t . The response variable y is generated by equation (4.8), where γ_0 , γ_1 , δ_0 and δ_1 are predetermined values.

We run 1000 replications for each parameter combination of (τ, c, δ_1) , where τ ranges from 0 to 1, c from 0 to $-T$, and δ_1 is chosen between 0 and $\frac{1}{10}\gamma_1$.

The following table represents the parameters we choose for the simulation.

Table 4.1: Simulation Parameters

T	1000	γ_0	1
σ_1	0.1	γ_1	2
σ_2	1	δ_0	0.1

The mean and standard deviations of the 1000 estimations along with the true values of each coefficient is presented in the tables in next page.

Table 4.2 and Table 4.3 represent the results of the intercept and slope estimation, respectively, with their true values in the second column. Different values of δ_1 provide us with different true models. By assigning 0 to it, we obtain an OLS setting. Notice that in this case the true value of slope does not vary across the quantiles. The value of c controls the autocorrelation structure of the independent variable X_t , with the first order autocorrelation $\rho_1 = 1 + \frac{c}{T}$. Particularly, when $c = -T = -1000$, X_t is an *i.i.d.* sequence. When c gets greater, the autocorrelation becomes stronger. We require $c \leq 0$, and when $c = 0$, time series X_t is not stationary. We choose $c = 0, -500$, and -1000 in our simulation. Finally, the quantiles τ vary from 0.1 to 0.9, with an interval of 0.1, so that we can see the performance of the estimation across different quantiles.

The results in the table first confirm our conclusion of consistency of the estimates. From the table we can see that for all of the parameter combinations, the mean of the estimates is close to its true value, with an acceptably small standard deviation. The estimation for quantiles around the median is slightly better than the tail quantiles. This can also be seen from the standard deviations of the estimates. For both the intercept and slope estimates, as the quantile increases, the standard deviation gradually decreases for the lower quantiles, after passing the median, it then

Table 4.2: Simulation Results-Intercept

$\delta_1 = 0$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	0.8718	0.8766(0.0579)	0.8716(0.0263)	0.8722(0.0295)
0.2	0.9158	0.9175(0.0500)	0.9155(0.0224)	0.9163(0.0232)
0.3	0.9476	0.9488(0.0474)	0.9476(0.0206)	0.9480(0.0220)
0.4	0.9747	0.9761(0.0455)	0.9746(0.0198)	0.9750(0.0214)
0.5	1.0000	1.0014(0.0437)	1.0000(0.0201)	1.0004(0.0213)
0.6	1.0253	1.0272(0.0477)	1.0250(0.0200)	1.0262(0.0208)
0.7	1.0524	1.0531(0.0487)	1.0518(0.0208)	1.0527(0.0223)
0.8	1.0842	1.0847(0.0530)	1.0829(0.0227)	1.0847(0.0232)
0.9	1.1282	1.1269(0.0650)	1.1269(0.0268)	1.1297(0.0278)
$\delta_1 = 0.2$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	0.8718	0.8748(0.0619)	0.8698(0.0533)	0.8726(0.0457)
0.2	0.9158	0.9179(0.0525)	0.9158(0.0445)	0.9158(0.0381)
0.3	0.9476	0.9490(0.0477)	0.9467(0.0404)	0.9476(0.0353)
0.4	0.9747	0.9738(0.0443)	0.9748(0.0391)	0.9757(0.0344)
0.5	1.0000	0.9979(0.0442)	0.9990(0.0378)	1.0009(0.0346)
0.6	1.0253	1.0245(0.0446)	1.0249(0.0387)	1.0265(0.0350)
0.7	1.0524	1.0512(0.0465)	1.0523(0.0402)	1.0533(0.0365)
0.8	1.0842	1.0837(0.0497)	1.0849(0.0435)	1.0835(0.0398)
0.9	1.1282	1.1282(0.0616)	1.1292(0.0521)	1.1268(0.0484)

increases. This phenomenon is consistent with the form of the limiting distributions of the estimations, since the constant $\tau(1 - \tau)$ is maximized at $\tau = 0.5$.

For most of τ and δ_1 combinations, the sample standard deviations increase as c increases to zero. This is also expected. We know that when c gets closer to 0, the autocorrelation of X_t gets bigger, so the distribution of $[\int \bar{J}_c \bar{J}'_c]^{-1}$ in Theorem 1, which is the counterpart of the matrix $(XX')^{-1}$ in OLS, is more spread. Therefore the convergence rate of the estimation to the true value is slower when c is closer to zero. Notice that when $c = 0$, which is the case that X_t is not stationary, the

Table 4.3: Simulation Results-Slope

$\delta_1 = 0$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	2	1.9996(0.0536)	2.0009(0.0456)	1.9998(0.0546)
0.2	2	1.9999(0.0467)	2.0009(0.0389)	1.9995(0.0437)
0.3	2	1.9999(0.0432)	2.0002(0.0362)	1.9994(0.0413)
0.4	2	1.9999(0.0415)	2.0004(0.0346)	1.9995(0.0402)
0.5	2	1.9999(0.0408)	2.0004(0.0351)	1.9992(0.0398)
0.6	2	1.9998(0.0445)	2.0008(0.0348)	1.9984(0.0389)
0.7	2	2.0000(0.0444)	2.0012(0.0363)	1.9995(0.0418)
0.8	2	2.0000(0.0483)	2.0026(0.0391)	1.9991(0.0435)
0.9	2	2.0001(0.0598)	2.0028(0.0465)	1.9970(0.0524)
$\delta_1 = 0.2$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	1.7437	1.7399(0.1264)	1.7478(0.0970)	1.7433(0.0993)
0.2	1.8317	1.8333(0.1103)	1.8318(0.0809)	1.8332(0.0824)
0.3	1.8951	1.8973(0.0988)	1.8973(0.0734)	1.8957(0.0771)
0.4	1.9493	1.9524(0.0949)	1.9491(0.0720)	1.9471(0.0752)
0.5	2.0000	2.0040(0.0927)	2.0015(0.0692)	1.9978(0.0756)
0.6	2.0507	2.0515(0.0926)	2.0509(0.0715)	2.0481(0.0758)
0.7	2.1049	2.1078(0.0955)	2.1042(0.0741)	2.1029(0.0787)
0.8	2.1683	2.1719(0.1056)	2.1663(0.0800)	2.1697(0.0861)
0.9	2.2563	2.2612(0.1255)	2.2538(0.0961)	2.2591(0.1035)

estimation is still consistent.

Now let's compare the case $\delta_1 = 0$ with $\delta_1 = 0.2$. We know that in the first case, variable X_t is independent with the error term, so that the influence of X_t on Y_t remains the same for different quantiles. Reflected in the model, we can see that the true values for the slope do not change. In this case, OLS is a better option for us, and we need to deal with over-fitting issues with QR model. The first panel of Table 4.3 shows us the performance of QR in this case. We can see that for all of the quantiles, the estimates of the slope is close to the constant true value

2. By comparing the difference among themselves with respect to their standard deviations, we can see we won't reject the null hypothesis $\beta_{\tau_1} = \beta_{\tau_2} = \dots = \beta_{\tau_9}$. This means that if we apply QR in the case where OLS is a proper model, it still gives us consistent results, and based on its results we eventually turn to the right model.

Next, let us consider the performance of OLS model under quantile regression settings. With other parameter values in Table 4.1, we choose $\delta_1 = 0.4$ to make the true model further away from OLS settings. Again, we run 1000 replications for each parameter combination of (τ, c) , and the candidate values of (τ, c) are the same as before. Based on the results, we generate the following plots.

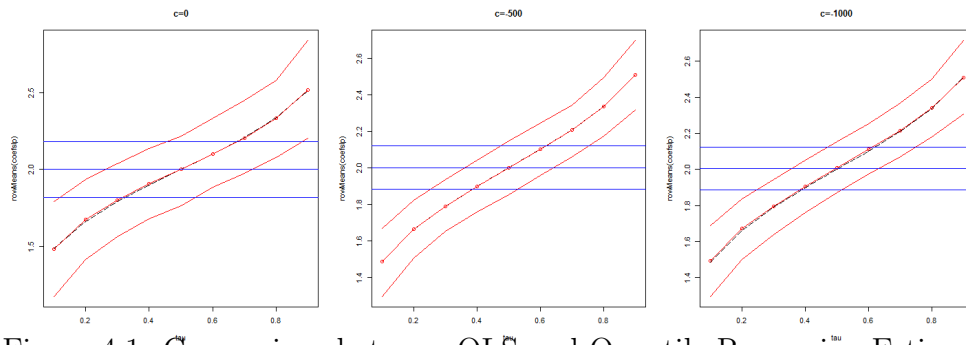


Figure 4.1: Comparison between OLS and Quantile Regression Estimates

The above plots compare the 95% confidence intervals of the slope estimations of OLS and QR models. The red lines and horizontal blue lines represent the confidence intervals under quantile regression models and OLS model, respectively. The upper and lower lines are the upper and lower bounds of the interval, while the middle lines are the locations of the mean of the 1000 estimates. The bounds of the confidence interval is computed as the 2.5th and 97.5th empirical percentile of the estimates. The black dashed line marks the true value of the slope across different quantiles. The plots are for the cases of $c = 0$, $c = -500$ and $c = -1000$ from left to right, respectively.

First, notice that the patterns are similar across different value assignments of c , which means that our conclusion is robust. However, when $c = 0$, the confidence interval is wider than other cases, which is consistent with the above conclusion that the limiting distribution is more spread when the independent variable is not stationary. Also, we can see that the mean of the estimates are close to the true mean, which confirms our consistency conclusion.

Since $\delta_1 = 0.4 > 0$, the true value along with the estimates of the slope of QR model is increasing as τ becomes greater. This means that the influence of X_t is greater on the larger values of Y_t , while OLS model cannot capture this phenomenon. We can see that the OLS confidence interval crosses with QR intervals at the place around median, while they stagger each other at the tails. This shows that OLS estimates only convey the average effect to the whole distribution. From the plots we can see that at the tails, the confidence intervals do not overlap, which means that the influence in the tails is significantly different from the average influence. Therefore we can draw the conclusion that under such circumstance, OLS is not adequate to provide a whole picture of the relationship between the two variables.

Finally, let's examine how far the limiting distribution is from a normal distribution. We know that the limiting distribution is not normal unless X_t is not autocorrelated. However, it is not a regular distribution as well. In order to perform inference analysis for the coefficient estimates, we need at least an approximate normality. We check the histogram and qq-plots of the 1000 estimations for each combination of the parameter (τ, c) . The following plots shows the results for $c = -500$ and $\delta_1 = 0.2$. The other combinations generate similar results.

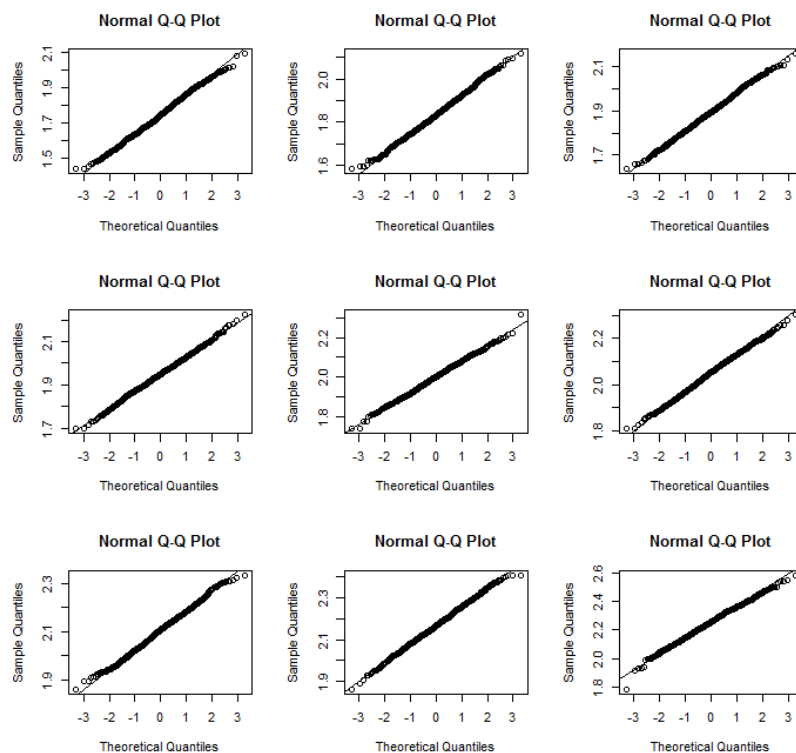


Figure 4.2: QQ Plots of Slope Estimation

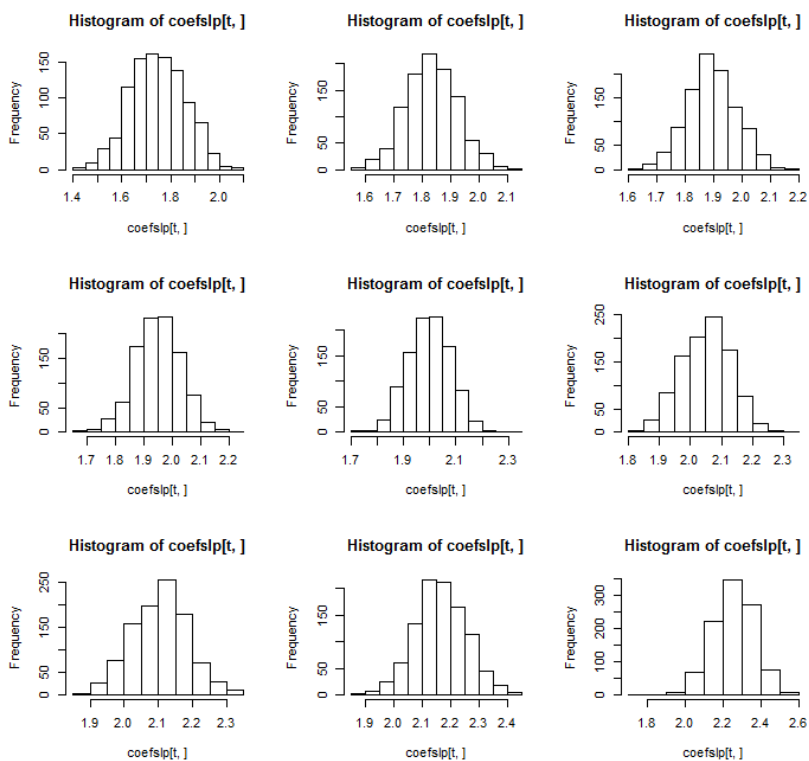


Figure 4.3: Histograms of Slope Estimation

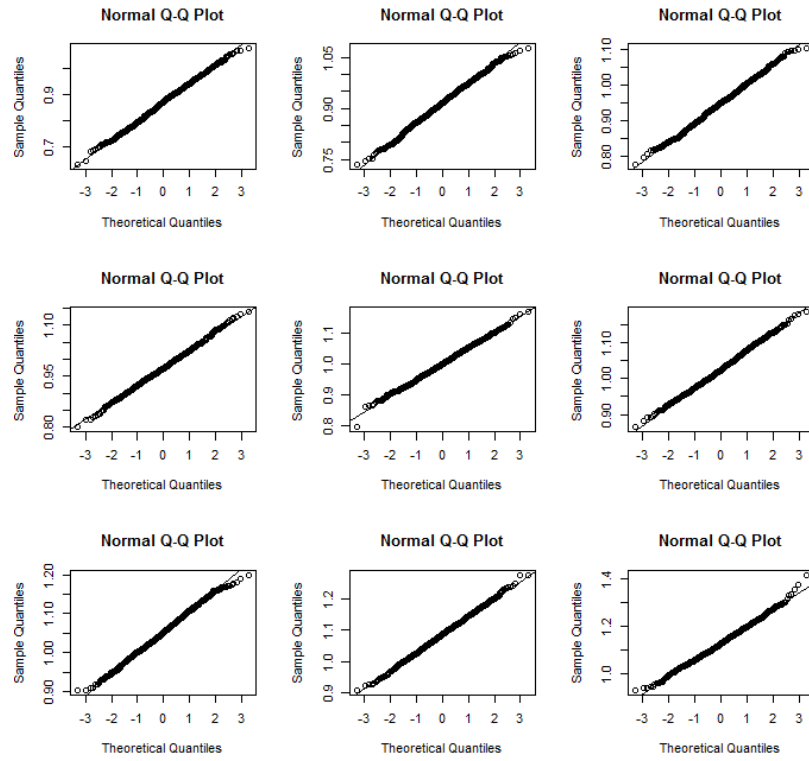


Figure 4.4: QQ Plots of Intercept Estimation

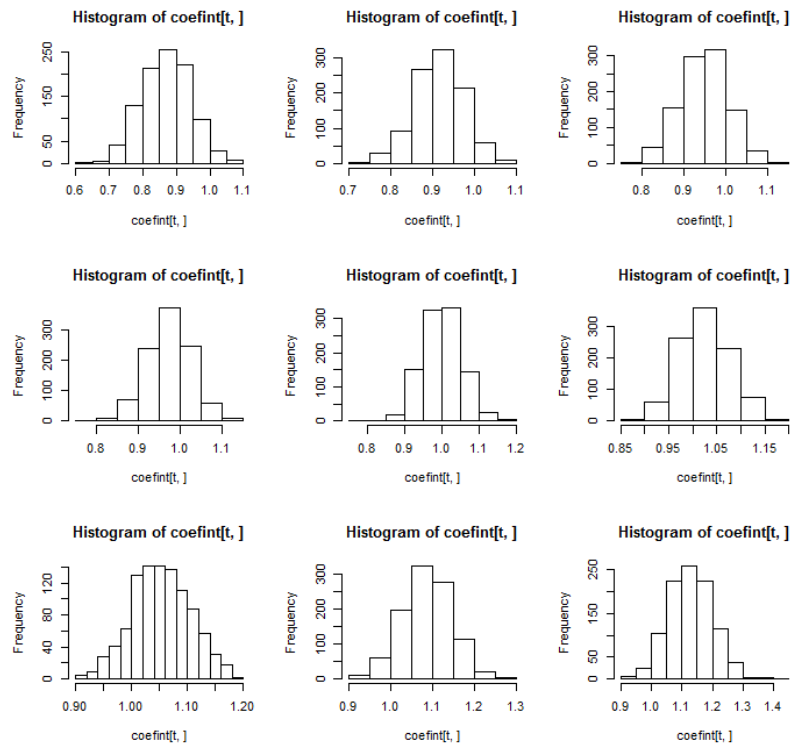


Figure 4.5: Histograms of Intercept Estimation

From the above plots we can see that the histograms show a normal pattern of the estimates, which is bell-shaped, single-peaked and with thin tails. This observation is also supported by the QQ plots. The dots are roughly linear though there is a little deviation at the tails. We can draw the conclusion that the limiting distributions is close to a normal distribution.

Since normality is served for statistical inference, we also compared the 'empirical' confidence interval with 'theoretical' confidence interval. The 'empirical' interval is calculated as above, the $\frac{\alpha}{2}$ th and $(1 - \frac{\alpha}{2})$ th empirical quantile of the 1000 simulation results, while the 'theoretical' one is computed under the assumption that the distribution is normal, *i.e.*

$$\bar{\hat{\beta}}_i \pm z_{1-\frac{\alpha}{2}} s_{\hat{\beta}_i}, \quad (4.12)$$

where $\bar{\hat{\beta}}_i$ is the sample mean of the simulation estimates, $s_{\hat{\beta}_i}$ is the sample standard deviation, and $z_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ th quantile of the standard normal distribution.

The following plots show the results of the comparison of the 95% confidence interval.

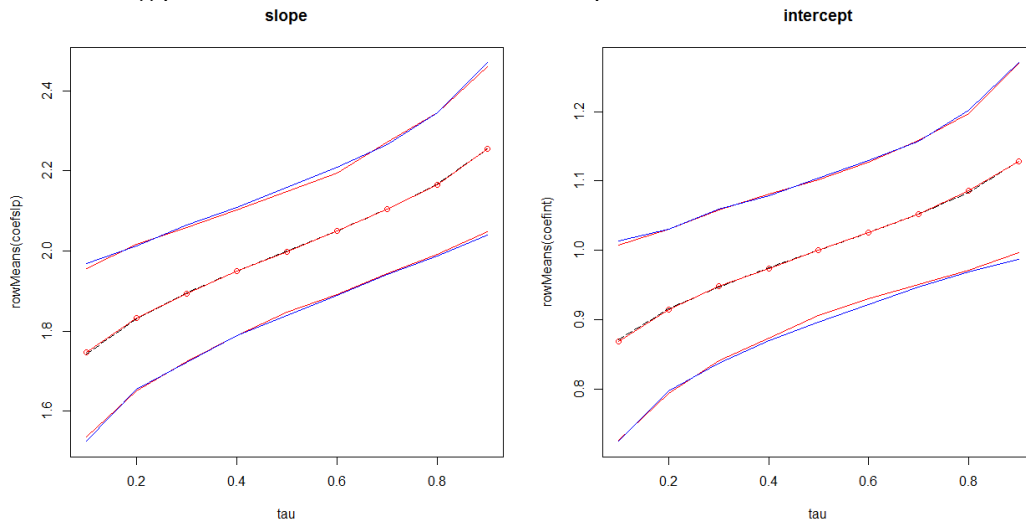


Figure 4.6: Comparison of Confidence Intervals

The middle lines are the average estimates along with the true values, the upper and lower red lines are the 'empirical' confidence intervals, and the upper and lower

blue lines are the 'theoretical' ones. From the plots we can see that for both slope and intercept, it does not induce much error if we assume normality when we give confidence interval estimation.

Next, we proceed further from the normal error term set up. Normal error is a standard assumption in an OLS model, however, when the noise no longer has such nice property, the estimations in OLS, though remain to be consistent, may not be efficient any more. Nevertheless, in our model, the assumption we put on the error term is martingale difference sequence, which allows us to explore a wide range of distribution families.

Now let us consider a skewed distribution - χ^2 distribution. Assume u_t in equation (4.7) follows a χ^2 distribution. To guarantee the noise term martingale difference sequence, we shift the distribution so that the mean is zero. To keep consistent with previous example, we also scale the distribution so that the standard deviation of the error term stays the same. We keep the linear dependence between x_t and ϵ_t as in (4.8). The fixed parameters are the same as Table 4.1, and the parameter combination of (τ, c, δ_1) are also kept consistent with previous study. The results we present here are obtained when u_t follows a χ^2 distribution with degree of freedom 4, which has skewness equal to $\sqrt{2}$. We also try different degrees, the results are similar.

Table 4.4 and 4.5, similar to table 4.2 and 4.3, show the results of simulation with skewed noise.

Table 4.4: Simulation Results for Skewed Noise-Intercept

$\delta_1 = 0$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	0.8962	0.8957(0.0272)	0.8962(0.0114)	0.8960(0.0106)
0.2	0.9169	0.9164(0.0319)	0.9170(0.0130)	0.9166(0.0123)
0.3	0.9362	0.9356(0.0363)	0.9362(0.0153)	0.9359(0.0140)
0.4	0.9559	0.9567(0.0421)	0.9553(0.0173)	0.9559(0.0157)
0.5	0.9773	0.9775(0.0475)	0.9773(0.0195)	0.9777(0.0174)
0.6	1.0016	1.0008(0.0524)	1.0020(0.0226)	1.0026(0.0207)
0.7	1.0311	1.0323(0.0635)	1.0308(0.0263)	1.0318(0.0245)
0.8	1.0703	1.0751(0.0808)	1.0701(0.0326)	1.0710(0.0293)
0.9	1.1336	1.1371(0.1146)	1.1317(0.0452)	1.1346(0.0417)
$\delta_1 = 0.2$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	0.8962	0.8962(0.0405)	0.8959(0.0226)	0.8952(0.0197)
0.2	0.9169	0.9156(0.0468)	0.9169(0.0249)	0.9159(0.0222)
0.3	0.9362	0.9342(0.0535)	0.9365(0.0279)	0.9346(0.0245)
0.4	0.9559	0.9556(0.0605)	0.9571(0.0307)	0.9551(0.0268)
0.5	0.9773	0.9779(0.0684)	0.9788(0.0346)	0.9753(0.0304)
0.6	1.0016	1.0015(0.0777)	1.0024(0.0390)	1.0010(0.0362)
0.7	1.0311	1.0334(0.0927)	1.0326(0.0451)	1.0297(0.0436)
0.8	1.0703	1.0756(0.1114)	1.0722(0.0577)	1.0706(0.0543)
0.9	1.1336	1.1362(0.1639)	1.1364(0.0852)	1.1365(0.0768)

Table 4.4 and 4.5 present the simulation results for intercept and slope estimates, respectively. From the results we can see that, first, our estimation is still consistent when we switch to a skewed noise set up. The estimations are close to the true values of the parameters, and the standard deviations of the estimates are relatively small. What is more, in the case when $\delta_1 = 0$, in which the error term is independent of the explanatory variable, the estimates of the slope remain stable across quantiles. This shows that in the case where OLS is actually preferred to quantile regression model, we can easily discern this misuse by simply comparing the results. Also,

Table 4.5: Simulation Results for Skewed Noise-Intercept

$\delta_1 = 0$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	2	2.0000(0.0212)	2.0001(0.0179)	2.0007(0.0214)
0.2	2	2.0001(0.0253)	1.9998(0.0201)	2.0009(0.0248)
0.3	2	2.0001(0.0286)	2.0001(0.0238)	2.0008(0.0283)
0.4	2	2.0000(0.0331)	2.0011(0.0272)	2.0003(0.0313)
0.5	2	2.0000(0.0378)	2.0001(0.0308)	1.9994(0.0348)
0.6	2	2.0001(0.0417)	1.9996(0.0354)	1.9983(0.0414)
0.7	2	1.9999(0.0500)	2.0005(0.0414)	1.9992(0.0493)
0.8	2	1.9996(0.0634)	2.0006(0.0517)	1.9997(0.0592)
0.9	2	1.9997(0.0905)	2.0032(0.0712)	1.9998(0.0830)
$\delta_1 = 0.2$				
		$c = 0$	$c = -500$	$c = -1000$
τ	True Value	Mean Est (SD)	Mean Est (SD)	Mean Est (SD)
0.1	1.7924	1.7926(0.0430)	1.7930(0.0421)	1.7948(0.0430)
0.2	1.8337	1.8348(0.0480)	1.8344(0.0459)	1.8364(0.0479)
0.3	1.8723	1.8739(0.0533)	1.8731(0.0521)	1.8759(0.0530)
0.4	1.9118	1.9122(0.0589)	1.9103(0.0573)	1.9142(0.0578)
0.5	1.9545	1.9543(0.0653)	1.9524(0.0639)	1.9592(0.0654)
0.6	2.0032	2.0033(0.0731)	2.0027(0.0719)	2.0047(0.0774)
0.7	2.0621	2.0597(0.0854)	2.0601(0.0840)	2.0654(0.0926)
0.8	2.1406	2.1352(0.1106)	2.1375(0.1061)	2.1402(0.1159)
0.9	2.2672	2.2646(0.1533)	2.2617(0.1551)	2.2613(0.1667)

the changing pattern of the estimation standard error with respect to different c values, which imply different autocorrelation structure of X_t , stays the same as in the previous results.

Next, let us compare the results in table 4.4 and 4.5 with those in table 4.2 and 4.3. Both of the estimates are consistent, while in the skewed error term case, the standard deviation of the estimates no longer achieves its minimum value around median. We can see that almost in all cases, the standard deviation gradually increases as the quantile gets greater. This phenomenon is not surprising since χ^2

distribution is a right skewed distribution; it has a long tail in the right while no tail at all in the left. Therefore, as the quantile proceeds to the higher part, the errors vary a lot more significantly than in the lower tail case, thus introduce more variance in the estimation. One interesting thing here is that given other parameters fixed, when we change the value of the quantile, most of the standard deviations of the estimates are less than those in the normal error case. For example, let us compare the second panel of table 5 with that of table 3. We can see that in $c = 0$ case, eight of the standard deviations are less than the minimum standard deviation in the normal error case with the same parameter combination. In other two cases, there are still at least six of them. Such pattern firstly indicates that quantile regression is a robust method. What is more, when we deal with skewed data, the model presents an overall better performance at the cost of poorer accuracy in the tail. We know that in the right skewed case, most of the data are aggregates in the range covering the lower quantiles, therefore, if we can determine that this somewhat 'normal' range is more important to our study, then the sacrifice of the accuracy in the tail estimations is worthwhile so that we can obtain a better precision for lower quantile estimations.

This observation is especially interesting when we research with economic data. We know that the errors in economic data are relatively less regular, and in quite a few cases, there are some unexpected shock in it. Therefore, we would like to separate the influence of the outliers from that of the normal data. However, this is hard to implement since most of the economic data evolve over time, therefore the noise are hidden under the structure of autocorrelation. For example, in our model, the factor X_t evolves according to the AR model, as in (3.1) and (3.2). Therefore by directly looking at the magnitude of X_t values, we can hardly determine which range the error term beneath X_t falls into. By this argument, we cannot separate data into

difference circumstances. However, with quantile regression, at least we can see that it provides us with accurate estimation under less extreme cases.

Next, let us compare the performance of quantile regression with OLS model in this skewed noise case. Plot 4.7 is constructed in the same way as Plot 4.1. We change $\delta_1 = 0.4$ to make the true model deviated further away from OLS settings.

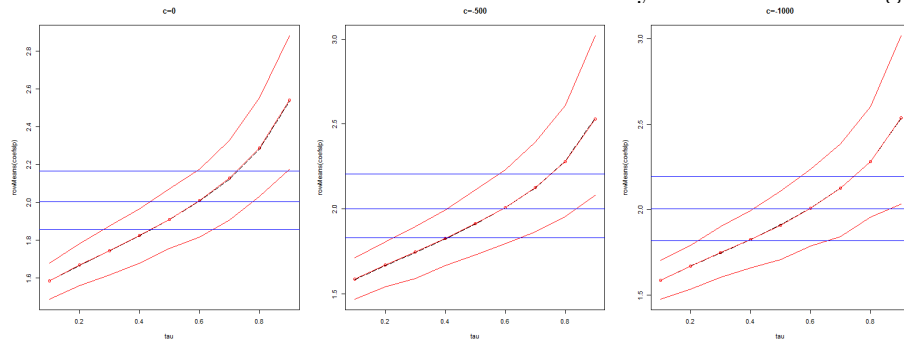


Figure 4.7: Comparison between OLS and Quantile Regression Estimates for Skewed Noise

Comparing Plot 4.7 with Plot 4.1, we can see that the difference of the patterns among the three cases is not as significant as that in Plot 1. This shows that the irregularity of the distribution of the noise to some extent absorbs the difference in the evolution dynamic of the data. Therefore, in the case where we have non-standard error terms, the quantile regression model is more robust to the autocorrelation structure of the time series data.

Another point that is worthwhile noticing is that, since the noise distribution is skewed, OLS estimate line no longer crosses QR estimate line at around $\tau = 0.5$. Since this distribution is skewed, the mean effect is drawn above by the long right tail so that they cross at around $\tau = 0.6$. This shows that the extreme values and outliers can impose considerable influence in OLS estimates. In the case where we study time series data, such extremeness is hard to explain since, as argued above, they are concealed in the autocorrelation structure of the data. Also, if we are more

interested in the lower quantiles' influence, where more data are clustered around, then OLS would give us a biased result. For example, if we assume the data above 80 percent quantile is influenced by an unexpected shock, and we want to check the performance of the data under normal circumstances, then by looking at the plots, we can see that the true average influence of X_t to Y_t in 'normal' range is close to the QR estimate around $\tau = 0.4$. However, this value lies outside of the OLS interval estimation for all of the three cases. So in the non-standard set up, quantile regression, as a more robust alternative of OLS, captures more properties hidden in the data.

We also checked the normality of the estimates. Again, we present here the histograms and qq-plots of the 1000 estimations given $c = -500$ and $\delta_1 = 0.2$. The other parameter combinations provide similar results.

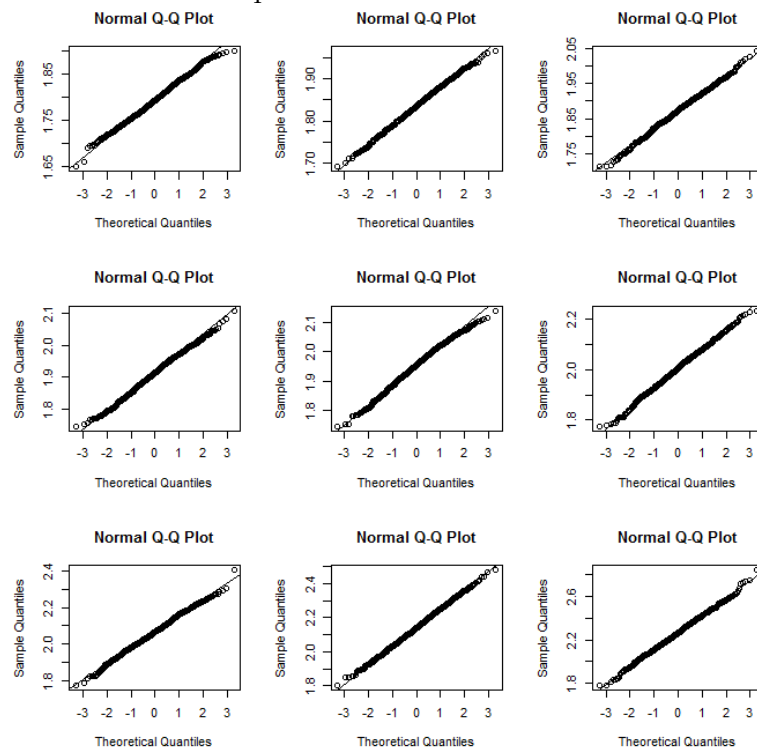


Figure 4.8: QQ Plots of Slope Estimation for Skewed Noise

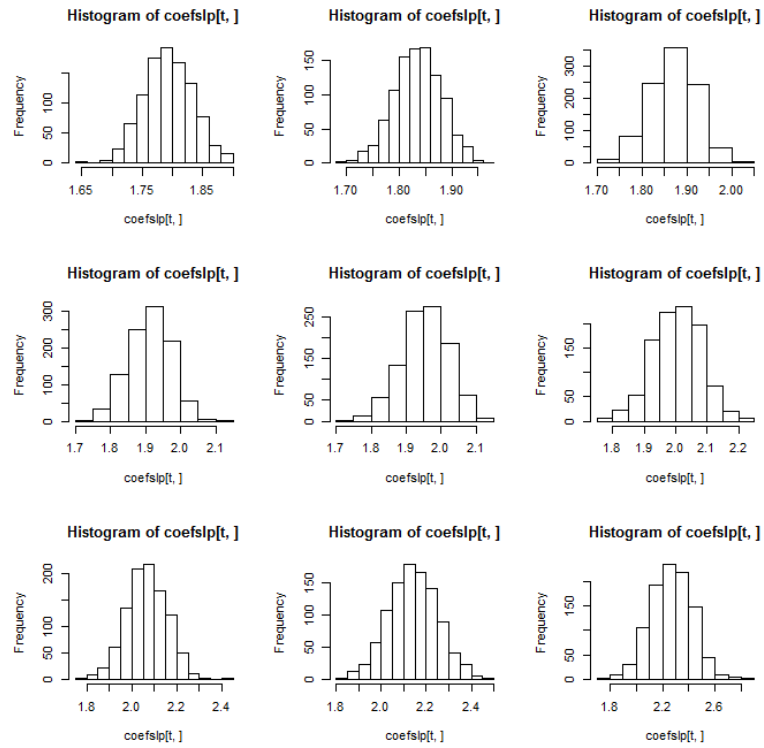


Figure 4.9: Histograms of Slope Estimation for Skewed Noise

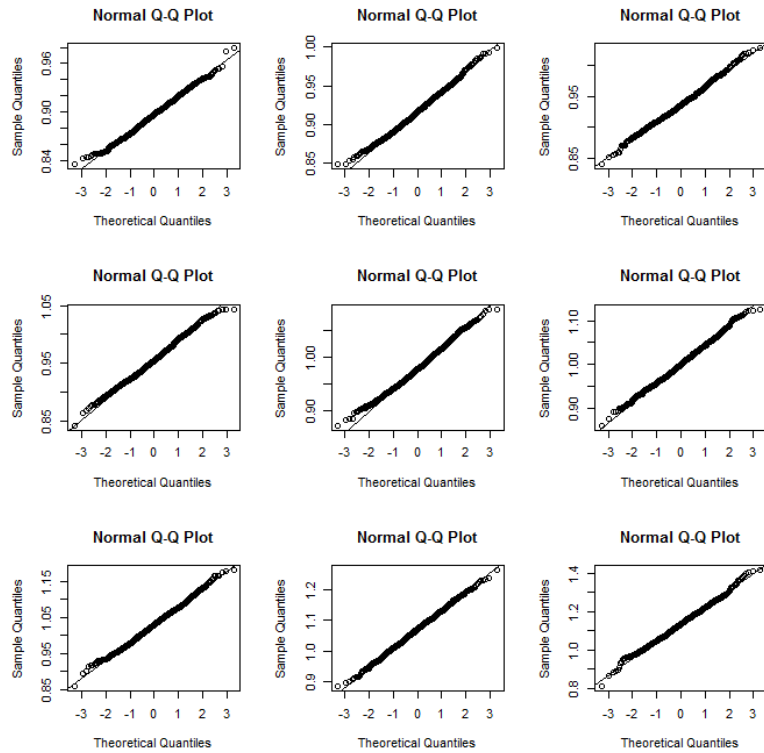


Figure 4.10: QQ Plots of Intercept Estimation for Skewed Noise

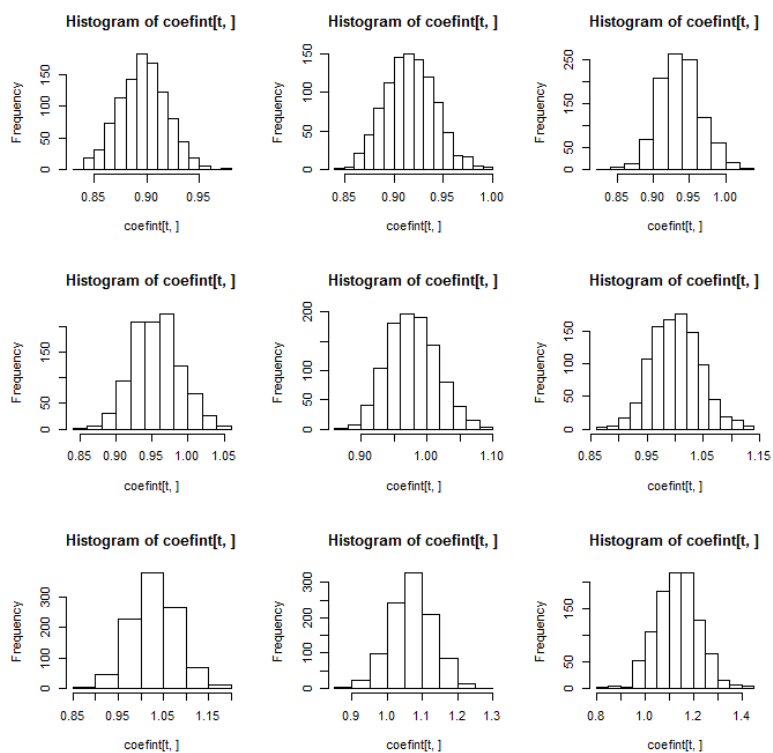


Figure 4.11: Histograms of Intercept Estimation for Skewed Noise

The above plots show that in the irregular case, the QR estimates asymptotic distributions are still close to normal. We also compare the confidence intervals constructed by different ways and obtained the following plot which is similar to Plot 5 above.

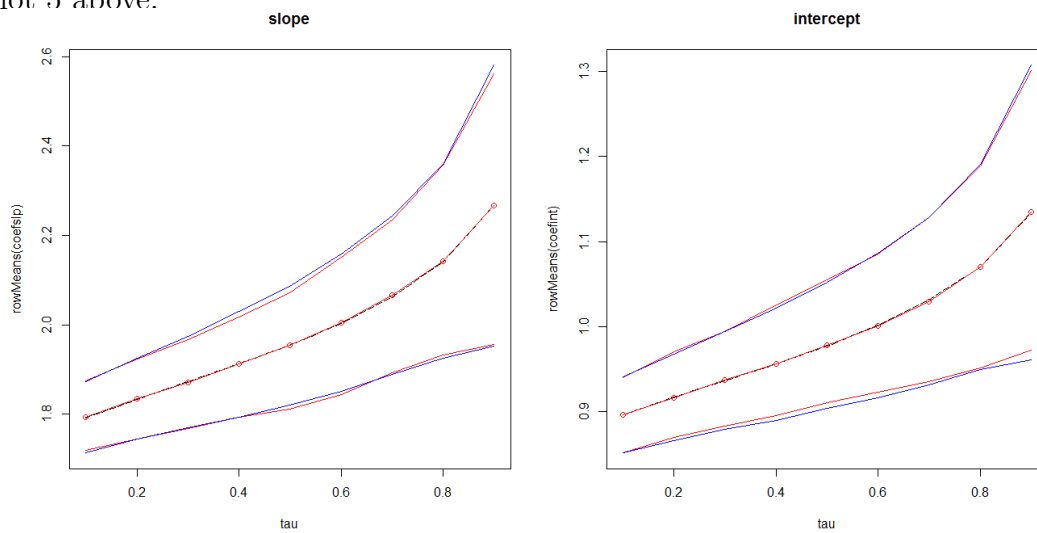


Figure 4.12: Comparison of Confidence Intervals for Skewed Noise

The plot shows that the "empirical" confidence interval is close to the one obtained by (4.12) for all of the quantiles. Therefore, in practice, we can use a normal approximation to perform inference to our model.

In this chapter, we further examine our model by simulation results. We show that under linear dependency between the error term and the influence factor X_t , quantile regression generates consistent estimates. Compared with OLS, quantile regression is more robust and captures more structure of the data. Also, normal distribution gives a satisfactory approximation of the asymptotic distribution of the estimates. These conclusions still hold when we deviate from the standard normal error term set up.

Chapter 5

Empirical Results

We use two sets of data to explore the application of our methodology in different areas.

5.1 Equity Risk Premium Forecast

5.1.1 Data Description

In this analysis, we use quarterly data from Welch and Goyal (2008), who provide a detailed description of the data sources. The reason that we choose this data set is due to the fact that it is widely employed by the literature, such as Rapach, Strauss, and Zhou (2010) and Meligkotsidou et al. (2012). In the analysis below, we compare our method with Meligkotsidou's, therefore we apply the same data set so that the results are more convincing. What is more, based on the literature, this set of variables have financial theoretical basis, which is also the reason why it is so widely employed.

The equity risk premium, which is the forecasted variable, is measured as the difference between the continuously compounded S&P500 returns, including dividends,

and the Treasury Bill rate, which is served as risk-free rate. The following 15 selected predictors are consistent with Zhou et al. (2010) and Meligkotsidou et al. (2012).

- *Earningspriceratio(log), E/P*: Difference between the log of earnings on the S&P 500 index and the log of stock prices, where earnings are measured using a one-year moving sum.
- *Dividendpayoutratio(log), D/E*: Difference between the log of dividends and the log of earnings.
- *Stockvariance, SVAR*: Sum of squared daily returns on the S&P 500 index.
- *Book – to – marketratio, B/M*: Ratio of book value to market value for the Dow Jones Industrial Average.
- *Netequityexpansion, NTIS*: Ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
- *Treasurybillrate, TBL*: Interest rate on a three-month Treasury bill (secondary market).
- *Long – termyield, LTY*: Long-term government bond yield.
- *Long – termreturn, LTR*: Return on long-term government bonds.
- *Termspread, TMS*: Difference between the long-term yield and the Treasury bill rate.
- *Defaultyieldspread, DFY*: Difference between BAA- and AAA-rated corporate bond yields.

- *Default returns spread, DFR*: Difference between long-term corporate bond and long-term government bond returns.
- *Inflation, INFL*: Calculated from the CPI (all urban consumers); following Welch and Goyal (2008), since inflation rate data are released in the following month, we use $x_{i,t1}$ in Equation (1) for inflation.
- *Investment-to-capital ratio, I/K*: Ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the entire economy (Cochrane 1991).

The quarterly data span from 1947:1 to 2011:4. We divide this sample period into in-sample and out-of-sample sub-periods, and use an expanding window to do the forecast, which will be illustrated in the following subsections.

5.1.2 Predictive Model

We aim to compare the performances of three models here. OLS, fixed weight QR model proposed by Meligkotsidou et al. (2012), and our new specification.

For each of the predictor x_i whose predictive power is of interest, we assume it has a linear relationship with the equity risk premium. The OLS model is given by

$$r_{t+1} = \beta_0 + \beta_1 x_{t,i} + \epsilon_{t+1,i}, \quad (5.1)$$

where $\epsilon_{t+1,i}$ is the error term when variable x_i is selected as the predictor. The prediction of OLS is given by:

$$\hat{r}_{t+1,OLS} = \hat{\beta}_0 + \hat{\beta}_1 x_{t,i}. \quad (5.2)$$

The fixed weight quantile regression model also applies quantile regression method. The method proceeds as follows. Given a set of quantiles, $\{\tau_1, \tau_2, \dots, \tau_L\}$, for each individual predictor x_i , we first estimate the quantile regression parameters with each τ_l , i.e.

$$Q_{r_{t+1}|X_{t,i}}(\tau_l) = \beta_{0,l} + \beta_{1,l}x_{t,i} + \epsilon_{t+1,\tau_l}, \quad (5.3)$$

where $Q_{r_{t+1}|X_{t,i}}(\tau_l)$ is the τ_l 's quantile of the risk premium at time $t + 1$, and ϵ_{t+1,τ_l} is the error term whose τ_l th quantile is 0. The predicted τ_l 's quantile of r at time $t + 1$ is

$$\hat{Q}_{r_{t+1}|X_{t,i}}(\tau_l) = \hat{\beta}_{0,l} + \hat{\beta}_{1,l}x_{t,i}.$$

According to the fixed weight quantile regression method, the forecasted return based on the information of x_i at time $t + 1$ is calculated as a weighted summation of these L predicted quantiles. Let $\{w_1, w_2, \dots, w_L\}$ be a set of weights satisfying $\sum_{l=1}^L w_l = 1$, the forecasted return will be calculated as

$$\hat{r}_{t+1,fwQR} = \sum_{l=1}^L w_l \hat{Q}_{r_{t+1}|X_{t,i}}(\tau_l). \quad (5.4)$$

There are different ways to assign weights. They are classified into two categories, fixed weights and time varying weights. Fixed weights are preassigned values and do not change with time. Time varying weights change over time, and are usually based on the performance of the model. Here we compare our model with fixed weight method. To keep consistent with our model, we choose $L = 3$, and $\{w_1, w_2, w_3\} = \{\tau_1, \tau_3 - \tau_1, 1 - \tau_3\}$.

Our specification assumes the same relationship between individual predictor x_i and equity risk premium as in the fixed weight quantile regression settings. However, we

introduce Markov chain to model the market evolution as described in chapter 2. For each quantile, π_l , we use the same process to give the estimate of $Q_{r_{t+1}|X_{t,i}}(\pi_l)$. The estimation of equity risk premium based on x_i is given by:

$$\hat{r}_{t+1,MCQR} = \hat{P}_{1S_t} \hat{c}_1 \hat{Q}_{r_{t+1}|X_{t,i}}(\tau_1) + \hat{P}_{2S_t} \hat{c}_2 \hat{Q}_{r_{t+1}|X_{t,i}}(\tau_2) + \hat{P}_{3S_t} \hat{c}_3 \hat{Q}_{r_{t+1}|X_{t,i}}(\tau_3), \quad (5.5)$$

where \hat{P}_{ij} and \hat{c}_i are defined and estimated as described in chapter 2.

We use an expanding estimation window to generate the out-of-sample predictions. More specifically, we divide the total sample of T observations into the in-sample period and out-of-sample period. Denote the length of in-sample and out-of-sample as m and k , respectively. The forecast of the initial out-of-sample period $m + 1$ based on x_i is given by:

$$\hat{r}_{m+1,i} = f(\hat{\theta}_m, x_{m,i}),$$

where $f(\cdot, \cdot)$ is one of the three models we compare, OLS, fixed weight QR and our new specification. $\hat{\theta}_m$ is the estimated parameters based on the observations $\{r_{t+1}, x_{t,i}\}_{t=1}^{m-1}$. Particularly, in our specification, θ includes the coefficients for quantile regression, market state variables, transition probabilities and the constants that relate the quantiles with the expectations of the return at each state. In our forecast procedure, we update the estimation of all the parameters for different time periods. Moving forward, the next forecast is given by

$$\hat{r}_{m+2,i} = f(\hat{\theta}_{m+1}, x_{m+1,i}), \quad (5.6)$$

where $\hat{\theta}_{m+1}$ is estimated by $\{r_{t+1}, x_{t,i}\}_{t=1}^m$. Proceeding in this manner all the way to the end of the out-of-sample period, we obtain k out-of-sample forecasts of r . The reason why we employ this expanding window to forecast equity risk premium is

because it simulates the behavior of investors in real world. Investors observe and accumulate information over time.

Following Campbell and Thompson (2008) and Welch and Goyal(2008), the benchmark we choose as the baseline of the comparison is the historical average. We also use an expanding window for the benchmark, i.e., $\bar{r}_{t+1} = \frac{1}{t} \sum_{j=1}^t r_j$, for $t > m$.

In the empirical study, we choose to divide the sample period into 1947:1-1964:4 and 1965:1-2011:4 to be consistent with Meligkotsidou et al(2012). We generate the out-of-sample analysis based on each of the 15 predictors from Welch and Goyal(2008). We employ the combination approach proposed by Rapach, Strauss and Zhou (2009) to give the ultimate forecasts. The combination method utilizes the information of all the single-factor models. The intuition behind this method is that, the prediction from a single factor can be structural unstable and contains a considerable number of "false signals", however, by averaging these forecasts, the simplest type of combination, the results should be more reliable. As Rapach, Strauss and Zhou show in their 2009 paper, the combination method not only generates consistent significant out-of-sample gains, but the idea is also linked to real economy.

The method works in the following way. Suppose we have N individual predictors, $\{x_1, x_2, \dots, x_N\}$. We estimate the individual model by each of them. Therefore, at time t , there are N forecasts of the equity risk premium of time $t + 1$, denoted as $\{\hat{r}_{t+1,1}, \hat{r}_{t+1,2}, \dots, \hat{r}_{t+1,N}\}$. The forecast of the combination method is given by

$$\hat{r}_{t+1,c} = \sum_{i=1}^N \omega_{t,i} \hat{r}_{t+1,i}, \quad (5.7)$$

where $\{\omega\}_{i=1}^N$ are the weights of individual factors at time t . The combination methods can also be categorized into two classes based on the assigned weight values: fixed weights and time varying weights.

The fixed weight method assigns weights that does not change over time. It includes the simplest combinations: mean, median and trimmed mean. The mean combination is given by setting $\omega_{t,i} = \frac{1}{N}$ for $i = 1, 2, \dots, N$, the median combination by setting $\omega_{t,i} = 1$ for the median and 0 otherwise. For the trimmed mean combination method, we set $\omega_{t,i} = 0$ for the n smallest and n largest values, and $\omega_{t,i} = \frac{1}{N-2n}$ for the rest.

The time varying weight method assigns weights by the historical forecasting performance of the individual models. Therefore, in order to calculate these weights, we need to leave out a holdout period. Suppose we use the first k_0 observations of the out-of-sample period, and we want to forecast the risk premium at time $t = m + k_0 + 1$. One of the possible weight assignments, discount mean square prediction error (DMSPE), is proposed by Stock and Watson (2004), as

$$\omega_{m+k_0,i} = \frac{\phi_{m+k_0,i}^{-1}}{\sum_{j=1}^N \phi_{m+k_0,j}^{-1}}, \quad (5.8)$$

where

$$\phi_{m+k_0,j} = \sum_{s=m}^{m+k_0-1} \theta^{m+k_0-1-s} (r_{s+1} - \hat{r}_{s+1,j})^2$$

and θ is a discount factor with preassigned values. From the definition of weights, we can see that DMSPE assigns greater weights to models with better forecast performance, measured by the squared prediction error. As we moving forward, the holdout period becomes longer. For example, to predict risk premium at time $t + 1$, where $t \geq m + k_0$, the holdout period starts from time $m + 1$ to time $t - 1$. The discount factor θ is usually chosen to be smaller than or equal to 1. When $\theta = 1$, there is no discounting; when $\theta < 1$, we put more weights on the more recent time periods.

The results of fixed weight combination method is included in subsection 4.2.3, and we will consider time varying weight combination approach in the future study.

5.1.3 Empirical Results

In this subsection, we compare the predictive power of our specification, fixed weight QR, OLS and the benchmark: historical average. Therefore we use the out of sample criteria: out of sample R-square R_{os}^2 , $ENC - T$ statistics, utility difference ΔU and wealth gain ΔW . For quantile regressions, the quantiles we choose are $\{0.25, 0.5, 0.75\}$. The aim of this selection is to keep consistent with related research. What is more, in this application, we are more interested in the forecast, other than the relationship between variables. Therefore we do not need to examine tail behaviors. Choosing less extreme values enables us to have moderate sample size for all of the three states, so that the estimation is more reliable. We also try different quantiles, the results show that the conclusion is robust with respected to the selection of quantiles. Accordingly, the weights in the fixed weight QR methods are chosen as $\{0.25, 0.5, 0.25\}$. The statistics are calculated for both individual models and the combination forecast. Each individual model is based on the information of one of the fifteen predictors. We consider three ways to combine them: mean, median, and trimmed mean. The trim parameter is chosen as $2/15$, which means we abandon the largest and smallest predicted values among the individual predictions, therefore it is less sensitive to outliers.

Table 5.1 represents the results of R_{os}^2 for individual models. R_{os}^2 compares one model with the benchmark - historical average. Positive values indicate that the model outperforms the benchmark in the sense of mean squared prediction errors. From the results we can see that most of the variables give negative R_{os}^2 values. This

Table 5.1: Individual Model Results-Out of Sample R^2

	OLS	fwQR	MC QR		OLS	fwQR	MC QR
D/P	0.007	0.020	0.000	TMS	-0.030	-0.063	-0.145
D/Y	0.009	0.023	-0.009	NTIS	-0.021	-0.038	-0.022
E/P	-0.012	-0.013	-0.019	INFL	-0.008	-0.003	0.000
D/E	-0.015	0.001	-0.006	LTR	-0.011	-0.042	-0.163
B/M	-0.018	-0.028	-0.017	DFR	0.002	-0.006	-0.011
TBL	-0.027	-0.025	-0.102	SVAR	-0.111	-0.150	-0.263
DFY	-0.028	-0.060	-0.024	I/K	-0.028	-0.026	-0.088
LTY	-0.027	-0.024	-0.079				

pattern is consistent with the conclusion in Welch and Goyal (2008), the individual predictor is unable to consistently deliver superior out-of-sample performance over historical average.

Table 5.2 gives the values of $ENC - T$ statistics. The test is to show whether one model encompasses another. We conduct the test to compare our model with historical mean, OLS and fixed weight QR, respectively. The left panel of the table presents the results of comparing our model against other methods, significant results mean that the other method does not encompass our model, *i.e.*, our model contains useful information for forecast that is not represented in the other model. The right panel of the table reports the results of comparing another model with our methodology. The table also reports the significance level of the test.

From the results we can see that most of the tests, both ways, related to historical average are significant. This shows that although the results are negative in the sense of squared prediction error, neither model encompasses the other; our model still contains some useful information that historical average does not. This means that although the prediction of individual model does not improve prediction accuracy, the predictors still contain useful forecast information. However, the com-

Table 5.2: Individual Model Results: ENC-T

	ENC-T: MC QR against			ENC-T: Other models against MC QR		
	hstr mean	OLS	fw QR	hstr mean	OLS	fw QR
D/P	1.826**	0.599	0.427	1.751*	1.153	1.678*
D/Y	1.850**	0.101	0.095	2.030**	1.431*	1.965**
E/P	0.860	0.550	0.741	1.253	0.925	0.940
D/E	2.009**	1.845**	1.425*	2.295**	1.411*	1.744**
B/M	0.940	0.959	1.449*	2.008**	0.921	0.432
TBL	2.123**	-0.206	-0.293	3.657**	1.313*	3.378***
DFY	1.513*	1.625*	2.440***	2.260**	1.372*	0.365
LTY	2.000**	0.369	0.392	1.598*	3.125***	1.611*
TMS	1.938**	-1.448	-1.168	1.486*	2.284**	1.716*
NTIS	1.453*	1.360*	1.944**	2.676***	1.433*	0.669
INFL	2.043**	1.643*	1.671**	1.752**	1.158	1.376*
LTR	1.270	-0.157	-0.651	1.520**	2.927**	2.694***
DFR	1.583*	0.760	1.130	1.965**	1.473*	1.328*
SVAR	1.915**	-0.700	-0.418	1.114	1.112	0.976
I/K	2.194**	0.219	0.166	2.823***	2.271***	2.283***

parison with OLS and fixed weighted QR show that for individual predictors, our model contains similar information with OLS and fixed weighted QR. This result is consistent with Rapach, Strauss and Zhou (2009), where they show the information in individual models is limited.

Table 5.3 presents the results of the combination methods. First, from the table we can see that combination methods improve the forecast power of all the models. All of the three models give negative R_{os}^2 values in the individual predictor case. However, when we apply combination methods the values turn positive. The upper panel of the results shows that compared with historical return, OLS reduces about two to three percent of the mean squared prediction error, which is consistent with the figures in Rapach, Strauss and Zhou (2009). The fixed weighted QR performs similarly. This is because to keep consistent with our model, we only choose three quantiles. The performance gets better when we increase the number of quantiles in

Table 5.3: Combination Model Results

		mean	median	trimmed mean
out of sample R^2	OLS	0.029**	0.022**	0.027**
	fw QR	0.029**	0.023**	0.026**
	MC QR	0.060***	0.053***	0.060***
ENC-T: MC QR against	hstr mean	3.129***	3.228***	3.204***
	OLS	2.473***	2.520***	2.613***
	fw QR	2.423***	2.580***	2.595***
ENC-T: other models against MC QR	hstr mean	-0.419	-0.606	-0.554
	OLS	-0.308	-0.288	-0.372
	fw QR	-0.109	-0.010	-0.175
Utility Difference	hstr mean	4.984	4.028	4.779
	OLS	2.633	3.166	2.988
	fw QR	2.900	3.256	3.263
Wealth Difference	hstr mean	78.803	51.786	71.793
	OLS	62.635	46.318	60.613
	fw QR	43.795	31.784	46.575

the model until some threshold. However, the improvement is limited, and the value of R_{os}^2 is never greater than four percent. The third row shows that our specification reduces around six percent of the mean squared prediction error. These doubled R_{os}^2 values show that our method has greater predictive power than the other models. The $ENC - T$ statistic also gives us encouraging results. From the table we can see that the tests that compare our model against other methods are all significant. This means that none of the other models encompasses our model. Also, the third panel shows that the results of $ENC - T$ tests which compare other models against our method are all negative. Thus we cannot reject the null hypothesis that our model encompasses the other models. Putting the results together, we can draw the conclusion that our model contains "more useful information" than other models. Compared with historical average, our model contains the information of the predictors, so do OLS and fixed weight QR methods. However, the specific information

that our model includes are conditional distribution of equity risk premium, which is not included in OLS, and market evolution dynamics, which is not in the fixed weight QR set up. The results of both way of $ENC - T$ tests show that all of the above information are valuable in equity risk premium forecast.

The last two panels of the table show that the utility differences and the wealth gains are all positive. This shows that a risk-averse investor, who uses mean-variance framework to construct her portfolio, is expected to obtain higher utility if she applies our model to predict equity risk premium. The values are represented in percentage. On average, the utility of an investor increases by about four to five percent if she uses our model instead of simply calculating the historical mean. The investor can also enjoy a two to four percent utility gains if she turns from OLS or fixed weight QR to our model. This result gives our model practical values. It shows that better prediction leads to higher utility. In the market where an investor aims to maximize her utility, she tends to choose our model.

The wealth change of our model is also encouraging. Starting with one unit of wealth, by applying our model for 47 years, the investor can expect to earn 50-80 more dollars compared with historical average. She also earns 40-60 dollars more when she chooses our model over OLS. The wealth gain is still around forty dollars compared with fixed weight quantile regression model. Therefore selecting our model induces higher profit for investors.

The empirical results are consistent with the relative work in this area, and also show that our model outperforms OLS and fixed weight QR method when we choose the same quantile set. We also changed the parameters we use above to check the robustness of our model.

5.1.4 Robustness Check

We first try different quantile selections and different sample periods. We choose other two sets of quantiles $\{0.3, 0.5, 0.7\}$ and $\{0.2, 0.5, 0.8\}$. The two different out-of-sample periods are consistent with Rapach, Strauss, and Zhou(2010). Table 5.4 and Table 5.5 show the results.

Table 5.4: Robustness Check: Different Quantiles

		$\tau = 0.3, 0.5, 0.7$			$\tau = 0.2, 0.5, 0.8$		
	combination method	mean	median	trimmed mean	mean	median	trimmed mean
out of sample R^2	OLS	0.029**	0.022**	0.027**	0.029**	0.022**	0.027**
	fw QR	0.029**	0.022**	0.025**	0.028**	0.017**	0.024**
	MC QR	0.062***	0.070***	0.062***	0.053***	0.049***	0.051***
ENC-T: MC QR against	hstr mean	2.535***	3.241***	2.899***	3.083***	3.365***	2.632***
	OLS	1.962***	2.660***	2.479***	2.399***	2.559***	2.413***
	fw QR	1.922***	2.796***	2.615***	2.319***	2.823***	2.684***
ENCT: Other models against MC QR	hstr mean	1.008	-0.391	0.191	-0.614	-0.888	-0.674
	OLS	0.996	-0.347	0.249	-0.327	-0.371	-0.270
	fw QR	0.013	-0.099	0.416	-0.135	-0.338	-0.132
Utility Difference	hstr mean	4.674	3.726	4.439	4.509	3.353	4.129
	OLS	2.323	2.864	2.647	2.158	2.491	2.337
	fw QR	2.549	3.040	2.880	2.535	3.005	2.784
Wealth Difference	hstr mean	76.160	48.405	60.891	58.850	36.220	48.170
	OLS	59.992	42.937	49.711	42.682	30.752	36.990
	fw QR	39.619	28.815	34.795	24.086	21.315	24.708

From the tables we can see that our model performs robustly better than the other models. That the better performance of our model against other benchmarks is not confined to a certain selection of quantiles is quite encouraging and meaningful in real use. This shows that our model does not depend on how the states are defined, which may be considered differently. On the other hand, the robustness for different out sample periods shows that our model is not constrained to the length of the

Table 5.5: Robustness Check: Different Sample Periods

		OS: 1976:1-2011:4			OS:2000:1-2011:4		
	combination method	mean	median	trimmed mean	mean	median	trimmed mean
out of sample R^2	OLS	0.006	0.005	0.004	0.002	-0.011	-0.002
	fw QR	0.005	0.004	0.003	-0.015	-0.022	-0.017
	MC QR	0.022**	0.012*	0.021**	0.023**	0.003*	0.023**
ENC-T: MC QR against	hstr mean	2.071***	2.134***	1.818**	1.256	0.791**	1.723**
	OLS	2.007***	1.944**	1.960***	1.376*	1.116	1.589*
	fw QR	1.828**	1.995***	1.729**	1.624**	1.309*	1.214
ENCT: Other models against MC QR	hstr mean	0.506	0.820	0.431	-0.043	0.580	-0.093
	OLS	0.397	0.869	0.333	-0.208	0.128	-0.381
	fw QR	0.417	0.886	0.395	-0.562	-0.188	-0.678
Utility Difference	hstr mean	1.732	0.879	1.581	3.535	1.757	3.327
	OLS	1.544	0.970	1.653	3.523	3.399	3.986
	fw QR	1.701	0.926	1.703	5.535	4.653	5.536
Wealth Difference	hstr mean	12.331	4.240	10.609	0.439	0.257	0.463
	OLS	13.100	4.854	12.917	0.468	0.408	0.539
	fw QR	0.053	2.513	0.742	0.502	0.372	0.528

prediction.

The next change we make to the tests is to leave the 2008 financial crisis period out of the calculation. For example, when we calculate R_{os}^2 , we exclude the time period after 2008:4, in the summation, *i.e.*

$$R_{os}^2 = 1 - \frac{\sum_{s=1}^{2008:3} (y_{m+s} - \hat{y}_{m+s,A})^2}{\sum_{s=1}^{2008:3} (y_{m+s} - \bar{y}_{m+s})^2}.$$

Notice that the data after 2008:4 are still used for the parameter estimations. The reason we want to check this is to see if the model still performs well for both the routine days and the bad days. Table 5.6 shows the results of this comparison.

The left panel of the table repeats the results in Table 5.3, while the right panel represents the results with periods 2008:4-end left out. From the table we can see

Table 5.6: Robustness Check: Financial Crisis

		Original Model Results			Out of sample results with 2008:4-end left out		
	combination method	mean	median	trimmed mean	mean	median	trimmed mean
out of sample R^2	OLS	0.029**	0.022**	0.027**	0.035**	0.027**	0.031**
	fw QR	0.029**	0.023**	0.026**	0.032**	0.025**	0.028**
	MC QR	0.060***	0.053***	0.060***	0.061***	0.054***	0.059***
ENC-T: MC QR against	hstr mean	3.129***	3.228***	3.204***	3.008***	3.119***	2.653***
	OLS	2.473***	2.520***	2.613***	2.165***	2.259***	2.253***
	fw QR	2.423***	2.580***	2.595***	2.240***	2.467***	2.571***
ENCT: Other models against MC QR	hstr mean	-0.419	-0.606	-0.554	-0.366	-0.559	-0.430
	OLS	-0.308	-0.288	-0.372	-0.075	-0.088	-0.086
	fw QR	-0.109	-0.010	-0.175	0.035	0.077	0.027
Utility Difference	hstr mean	4.984	4.028	4.779	5.197	4.150	4.836
	OLS	2.633	3.166	2.988	2.523	3.075	2.786
	fw QR	2.900	3.256	3.263	2.923	3.303	3.199
Wealth Difference	hstr mean	78.803	51.786	71.793	51.199	32.590	43.040
	OLS	62.635	46.318	60.613	37.625	27.403	33.668
	fw QR	43.795	31.784	46.575	27.177	19.500	26.305

that most of the values of the statistics are comparable, except the wealth difference. This is expected since we calculate the accumulated wealth difference, and there are 11 less time period for the robustness calculations. Notice that the closeness of the two sets of the results shows that our model performs consistently well on both normal and bad days, since leaving out the crisis period does little influence to the results.

To further working on the performance of our model on different market states, we also test the out-of-sample forecast for different periods based on an economical variable. We choose real GDP growth to classify the market state. The procedure consists of two steps:

- Classify the whole sample period into three types of sub-sample periods: good, normal, and bad according to the value of real GDP growth, i.e., the economy is good if a value of real GDP growth for time t (quarter) is at the top 30% of all the values, bad if the value is in the bottom of 30%, and normal otherwise.
- Calculate goodness-of-fit for out-of-sample forecast for each of the three periods, i.e., R-square value as well as the test statistic.

The out of sample R^2 for different market states along with the whole sample period results are shown in the following Table 5.7.

Table 5.7: Robustness Check: GDP Indicator

		GDP State=1 (bad times)			GDP State=2 (mundane times)		
	combination method	mean	median	trimmed mean	mean	median	trimmed mean
R_{os}^2	OLS	0.020	0.014	0.019	0.041**	0.026*	0.031**
	fw QR	0.024*	0.016	0.020	0.035**	0.023*	0.029*
	MC QR	0.067**	0.072**	0.072**	0.066***	0.033**	0.050**
		GDP State=3 (good times)			whole sample results		
	combination method	mean	median	trimmed mean	mean	median	trimmed mean
R_{os}^2	OLS	0.035*	0.035**	0.038**	0.029**	0.022	0.027***
	fw QR	0.033*	0.038*	0.034*	0.029**	0.023	0.026***
	MC QR	0.039**	0.036***	0.046***	0.060***	0.053***	0.060***

The results are quite encouraging. The fact that the best performance appears in bad period is consistent with the findings of Rapach, Strauss, Zhou 2010, and furthermore our combination forecast is significantly better than the historical average for all the three periods. This is better than the combination forecast results because their forecast is not significantly better for normal period, see Table 5 in their paper.

This section we apply our model to Welch and Goyal's 15 factor data to forecast the equity risk premium. Different model selection criteria show that our model outperforms historical average, OLS and fixed weight QR model. What is more, results show that this conclusion is robust against the selection of quantile selection, the splitting of sample period and the status of the economy.

5.2 Fama-French Three Factor Model

5.2.1 Data Description

In the second application, we use monthly return data of thirty industry portfolios. With this data set, we apply Fama-French three factor model. We aim to examine the in sample relationship between the portfolio returns and Fama-French three factors. Therefore we concentrate on the in-sample relationship within this study. We also do some prediction analysis for the use of future research concerning portfolio construction.

The data is obtained from Kenneth R. French website. The sample period spans from 1927/07 to 2011/07. Each NYSE, AMEX and NASDAQ stock is assigned to an industry portfolio based on its four-digit SIC code. The portfolios are constructed at the end of June of each year. The returns are calculated in two ways: value-weighted and equal-weighted. Accordingly, we have the data of Fama French three factors for the same time period.

5.2.2 Fama-French Three Factor Model

The Fama-French three factor model is designed by Eugene Fama and Kenneth French as a generalization of the traditional asset pricing model, the Capital Asset

Pricing Model, CAPM. In CAPM, there is only one variable, beta, to predict the returns. However, the Fama-French model uses three factors as follow:

$$r = R_f + \beta_3(K_m - R_f) + b_s \cdot SMB + b_v \cdot HML + \alpha + \epsilon, \quad (5.9)$$

or

$$r - R_f = \beta_3(K_m - R_f) + b_s \cdot SMB + b_v \cdot HML + \alpha + \epsilon. \quad (5.10)$$

Where R_f is the risk-free return rate, K_m is the whole stock market return, SMB stands for 'small minus big', which measures the historic excess returns of small market capitalization over big ones, HML stands for 'high minus low', which measures the historic excess return difference between value stocks and growth stocks, and ϵ is the error term. Due to the inclusion of the latter two factors, the coefficient β_3 is analogous but not equal to the classical β in CAPM. Fama-French model is based on the observation that the small caps or high book-to-market ratio stocks tend to perform better, thus it includes the two factors, SMB and HML , to measure the historic excess returns of small caps over big caps and of value stocks over growth stocks.

Fama-French three factor model is widely used in forecasting stock returns. It explains over 90% of the diversified portfolio returns, while CAPM only explains about 70%. In this study, we also apply Fama French model for return forecast. However, to estimate the parameters, other than OLS which is generally used, we will try to combine it with the idea of quantile regression. We split the total sample into in sample period and out of sample period, and their lengths are denoted as m and k , respectively. We use the in sample data to examine the relationship between the three factors and the excess return, and out of sample data to check the predictive

power of both models.

In the OLS setting, it is assumed that there is a linear relationship between each factor and the excess return. The above error term ϵ is assumed to be normally distributed with mean 0 and a constant standard deviation, which leads to the normal assumption of the excess return. For the sake of brevity, denote excess return of portfolio i at time t as $R_{t,i}$, the Fama-French three factors as F_1 , F_2 , and F_3 , respectively. The in sample fitted values given by OLS method is

$$\hat{R}_{t,i} = \hat{\beta}_3 F_{1,t} + \hat{b}_s F_{2,t} + \hat{b}_v F_{3,t} + \hat{\alpha}, 1 \leq t \leq m, \quad (5.11)$$

where the estimation of the coefficients are obtained by minimizing the squared deviations. The coefficients are interpreted as the partial contribution of each factor to the expected value of the excess return.

Quantile regression, on the other hand, examines the relationship between the predicted factors and the quantiles of the excess return. Therefore we assume:

$$Q_{R_{t,i}|F}(\tau_l) = \beta_{3,\tau_l} F_{1,t} + b_{s,\tau_l} F_{2,t} + b_{v,\tau_l} F_{3,t} + \alpha_{\tau_l} + \epsilon_{\tau_l}, 1 \leq t \leq m, \quad (5.12)$$

where $Q_{R_{t,i}|F}(\tau_l)$ denotes the τ_l th quantile of the conditional distribution of portfolio excess return given the values of the three factors. ϵ_{τ_l} is the error term whose τ_l th quantile is assumed to be 0. We assume three states for the market, therefore $l \in \{1, 2, 3\}$ and the state of each time period S_t is identified. We consider two ways

of calculating the fitted values:

$$\hat{R}_{t,i,QR} = \begin{cases} \hat{Q}_{R_{t,i}|F}(\tau_1), & \text{when } S_t = 1 \\ \hat{Q}_{R_{t,i}|F}(\tau_2), & \text{when } S_t = 2 \\ \hat{Q}_{R_{t,i}|F}(\tau_3), & \text{when } S_t = 3 \end{cases}, \quad (5.13)$$

$$\hat{R}_{t,i,QR} = \begin{cases} \hat{c}_1 \hat{Q}_{R_{t,i}|F}(\tau_1), & \text{when } S_t = 1 \\ \hat{c}_2 \hat{Q}_{R_{t,i}|F}(\tau_2), & \text{when } S_t = 2 \\ \hat{c}_3 \hat{Q}_{R_{t,i}|F}(\tau_3), & \text{when } S_t = 3 \end{cases}. \quad (5.14)$$

The second way includes the conditional distribution specification at the three states. We try to see whether this specification induce significant fitness progress. As the empirical results show, these two ways give similar results.

For the consideration of future study, which may include portfolio construction, we also examine the predictive power of both models. To resemble the real world situation, we apply the 1-time period lag Fama-French model for both of OLS and our specification. The models are given by

$$R_{t+1,i} = \beta_3 F_{1,t} + b_s F_{2,t} + b_v F_{3,t} + \alpha + \epsilon_t, \quad (5.15)$$

for OLS, and

$$Q_{R_{t+1,i}|F}(\tau_l) = \beta_{3,\tau_l} F_{1,t} + b_{s,\tau_l} F_{2,t} + b_{v,\tau_l} F_{3,t} + \alpha_{\tau_l} + \epsilon_{\tau_l}, \quad (5.16)$$

for QR case.

The forecasted value of OLS is given by

$$\hat{R}_{t+1,i,OLS} = \hat{\beta}_3 F_{1,t} + \hat{b}_s F_{2,t} + \hat{b}_v F_{3,t} + \hat{\alpha}, \quad (5.17)$$

and the forecasted value of our specification is given by

$$\hat{R}_{t+1,i,QR} = \hat{c}_1 \hat{P}_{1,\hat{S}_t} \hat{Q}_{R_{t,i}|F}(\tau_1) + \hat{c}_2 \hat{P}_{2,\hat{S}_t} \hat{Q}_{R_{t,i}|F}(\tau_2) + \hat{c}_3 \hat{P}_{3,\hat{S}_t} \hat{Q}_{R_{t,i}|F}(\tau_3), \quad (5.18)$$

where \hat{S}_t is the estimation of market state at time t , and

$$\hat{Q}_{R_{t,i}|F}(\tau_l) = \hat{\alpha}_{\tau_l} + \hat{\beta}_{3,\tau_l} F_{1,t} + \hat{b}_{s,\tau_l} F_{2,t} + \hat{b}_{v,\tau_l} F_{3,t}.$$

The other parameters are estimated as described in chapter 2.

To test the robustness of the models, the prediction is performed in a fixed window format, this means that the parameters are estimated all based on the observations $\{Y_t, F_t\}_{t=1}^m$. The reason that we do not update the estimations of the parameters is that, the main purpose of the out-of-sample analysis is to examine whether the QR model based on the in-sample works better than the OLS. After fitting the in-sample data to the model, we obtain three regressions based on three different quantiles together with the transition matrix, which we use to characterize three possible scenarios for security returns as well as the dynamics of changes in these scenarios over time. Now the question is whether the model works well in practice. Therefore, we consider an out-of-sample and treat it as a random sample generated by the in-sample model. If the in-sample model fits the out-of-sample well, it provides evidence that the model works because it can reasonably reproduce the out-of-

sample; otherwise the model does not work well.

5.2.3 Difference Between Two models

The aim of this project is to examine the relationship between the three Fama French factors and portfolio excess returns. The main difference in the relationship assumption between the two set ups is that OLS assumes that the three factors only influence the mean of excess return distribution, which means that they only have a location shift contribution. While in quantile regression set up, we assume that the factors can influence each quantile of the conditional distribution differently, therefore, in addition to the location, the factors are also allowed to impact the shape of the distributions.

We first need to see whether the three factors in our model contribute significantly differently to different quantiles of stock returns. To check the necessity of quantile regression method, we regress excess returns on the three Fama French factors, obtain the partial influences of each factor to different quantiles of returns, and compare those influences to the OLS estimates. For each portfolio, we run the quantile regression for nine different quantiles: $\{0.1, 0.2, \dots, 0.9\}$, along with OLS estimates.

To give the readers a clearer view, we present the results in the forms of plots instead of tables here. Following is the plots of the results of the model for portfolio 20, utility.

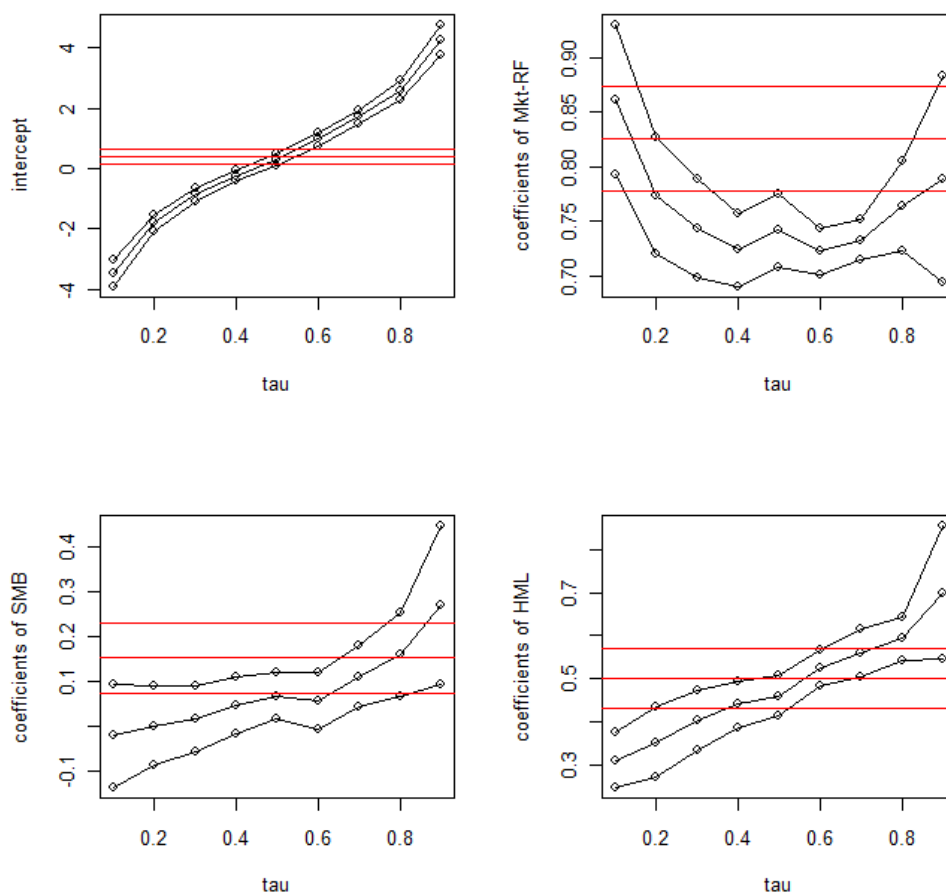


Figure 5.1: Comparison between OLS and Quantile Regression Estimates

In the plot, the upper and lower black lines indicate the interval estimations of coefficients for different quantiles, along with the point estimators denoted by the middle black line. The red lines are estimates obtained by OLS.

We can see from the plots that, OLS estimates are roughly the average of the quantile regression estimates. This shows that OLS coefficient estimates summarize the overall influence of each factor to the excess return. It is a more powerful estimation only when the location shift influence assumption is supported by the data. However, the patterns in the above plots lead to the opposite conclusion.

The intercept increases significantly according to quantiles as expected. The coefficient of market return minus risk free return, β_3 , behaves differently in the middle

quantiles. For both the upper and lower tails, the contribution is similar to the mean contribution, however, the influence to the middle part around median is significantly lower than the other parts. The coefficient of SMB increases by quantiles. An important feature of this plot is that the estimation changes its sign at around 30th quantile. This means that SMB is negatively related with lower level of excess return, while positively related with higher levels. OLS can not capture this relationship. HML coefficient exhibits similar pattern. The increase with quantiles shows that as the excess return increase, so does the contribution of SMB to the return. The results for other portfolios deliver similar patterns. Therefore, such results give us evidence and motivation to use quantile regression to predict portfolio returns.

5.2.4 In Sample Fitness

We remove the first 400 observations in order to be consistent with other studies in this area. The observations start from 1920s, and by removing the first 400 observations, the sample period starts from late 1960s. Therefore we still have observations for the last 50 years, which is long enough for our study. Furthermore, we also change the sample period for robustness check, and different sample periods give us similar results.

We split the remaining time period into two parts: sample period and prediction period. The former is for the in-sample analysis and the latter is for out-of-sample tests. We make the prediction period to be 2 years, which means that there are 598 observations left for in-sample analysis, and that is nearly 50 years, spanning from October 1959 till July 2009.

For the in-sample analysis, we fit the data to the model in such a way that the

value of the dependent variable is the portfolio return at time t , and the values of the independent variables are also the values of the factors at time t . There is no lag in time between the dependent and independent variables. This is because for the capital asset pricing model (CAPM) or Fama-French factor models, it specifies a contemporaneous relation between returns on a sector portfolio and factors.

For in-sample analysis, we use the observed values for the three factors at time t as the values of the independent variables, and the observed value on the return of one sector portfolio at time t as the value of the dependent variable.

We try symmetric quantiles $\tau = \{0.1, 0.5, 0.9\}$ first, but since the probability matrix is also roughly symmetric, therefore when we predict the future returns, the lower and upper quantiles cancel out, giving results similar with median. Median predictions are not significantly better than OLS, in which we use mean predictions. Therefore we end up with a set of asymmetric quantiles $\tau = \{0.05, 0.5, 0.9\}$.

The objective of the in-sample analysis is to estimate the model parameters, *i.e.*, 30 sets of the regression parameters for the OLS, one for each portfolio, and 3×30 sets of the regression parameters plus a 3×3 transition probability matrix. With the parameter estimates, we can calculate the fitted returns for both the OLS and the QR methods, and compare them with the actual returns.

We do not present the parameter estimations here due to space limitations. However, we can get similar conclusions as from the previous plots, so we will not iterate them here. One thing that needs to be noticed is that for most of the cases, the coefficients estimates are significantly different for different quantiles.

We use the two in sample criteria: mean squared error and mean absolute deviation to compare the fitness of the two models. The results are summarized in the following plot.

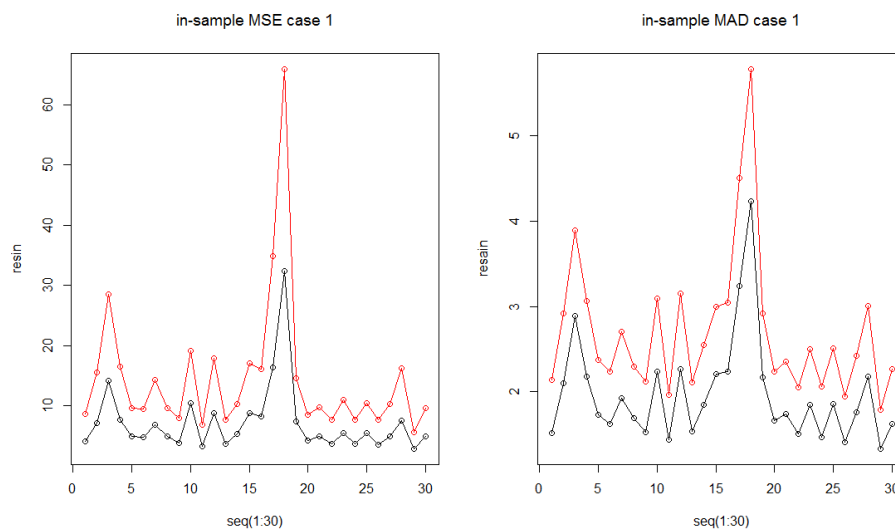


Figure 5.2: In Sample Analysis for French-Fama Model

The plots present the fitness of two models for 30 portfolios. The black lines and green lines are the residuals of the two quantile regression methods in (5.13) and (5.14), respectively, and the red lines are those of OLS. From the plots we can see clearly that QR fits the data better than OLS. Specifically, both the mean square error and the mean absolute deviation are less in the quantile regression model, and this is true for all of the 30 different portfolios. Also, the two quantile regression methods give really close results. However, examining in a finer scale, we can see that the specification with distribution assumption gives slightly better results than the one without distribution assumption.

5.2.5 One time period lag model

Portfolio construction is based on the prediction of future returns. Therefore, to check the predictive power of our specification, we need to resemble the real world situation. Here, we use 1 period lag model, *i.e.*, we rely on the value of three factors today to predict portfolio returns for tomorrow.

We first analyze the in-sample performance for this model. The fitted values are

calculated as described in Section 5.2.2.

One time period lag model is aimed to check the predictive power of the Markov chain quantile regression specification, therefore we need to employ our Markov chain framework into this analysis. As is indicated before, besides the parameters, we also estimate the states for each time period. If the return falls below 5th percentile, we assign state 1 to that time period. Returns above 90th percentile indicate state 3, and otherwise is state 2. We assume that $\{S_t\}$ follows a three-state first-order Markov chain. Therefore our first job of prediction is to estimate the transition probability matrix.

We assume that all industry markets evolutions are controlled by the same probability matrix. Altogether we have 598 observations for each sector of portfolios, and we have 30 sectors in total. Therefore we obtain 598×30 observations of states, thus 597×30 observations of transitions. We assume that all of the 30 sectors follow the same mechanism; therefore they share the same transition probability matrix. The estimates of transition probabilities are as follows:

Table 5.8: Estimation of Transition Probabilities

Probability Estimates				Standard Errors			
state	to state			state	to state		
	bad	median	good		bad	median	good
bad	0.0783	0.8297	0.0929	median	0.0020	0.0028	0.0022
median	0.0463	0.8589	0.0948	median	0.0016	0.0026	0.0022
good	0.0605	0.7842	0.1553	good	0.0018	0.0031	0.0027

From the above table we can see that all of the probabilities are significant. What is more, if we add each column and standardize, the marginal probabilities are roughly 0.05, 0.85, 0.1, which is exactly what we assume.

The following plots show the result about fitness.

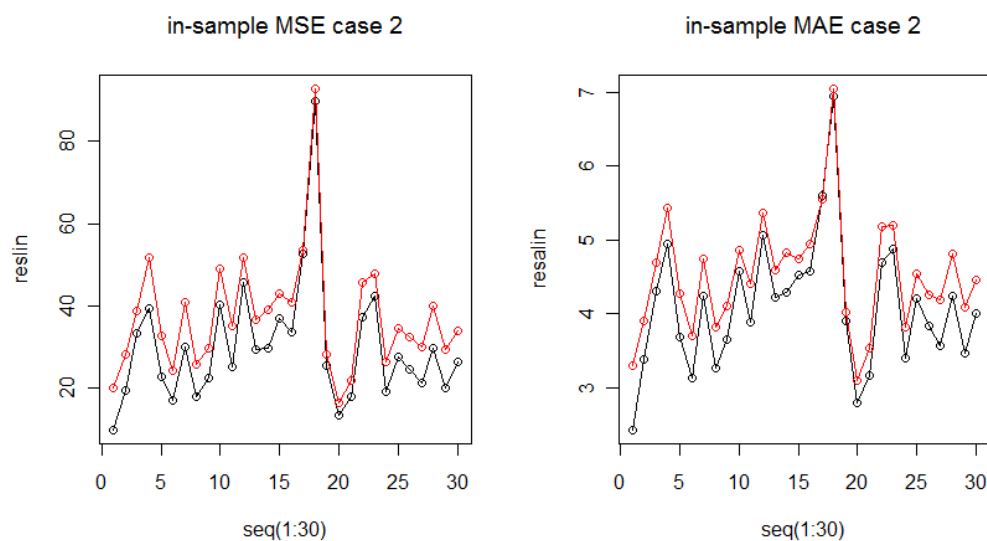


Figure 5.3: In Sample Analysis for French-Fama Mode: Case 2

From the plots we can see that quantile regression still outperforms OLS for in sample case. For each of the 30 sector of portfolios, quantile regression model fits the data better, giving both a less mean square error and a less mean absolute deviation.

In out-of-sample analysis, we use the parameter estimates obtained from in-sample analysis.

To predict the out-of-sample values, although we do not update parameter estimates throughout the procedure, we do assume we know previous day's information. For example, when we predict returns in month 660, we assume that we know all the information in month 659, including the three factors and the market state, even though we do not include it in the regression.

We compare mean squared/absolute prediction error and $ENC - T$ statistics for out of sample period.

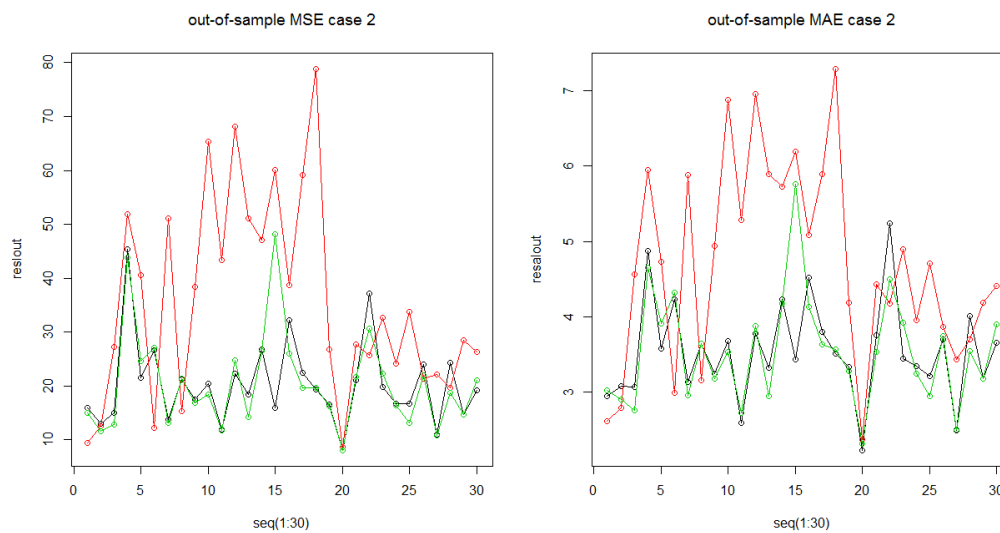


Figure 5.4: Out Sample Analysis for French-Fama Mode

From the plot we can see that overall, quantile regression has a stronger predictive power. However, for some of the sectors, such as sector 8 health, sector 22 service, OLS is still better than quantile regression.

The $ENC - T$ statistics are represented in the following table 5.9.

Table 5.9: ENC-T test statistic that compares MC QR with OLS

Industry	W/ Dist.	W/O Dist.	Industry	W/ Dist.	W/O Dist.	Industry	W/ Dist.	W/O Dist.
Food	1.258*	1.006	Clths	1.294*	1.290*	FabPr	0.887	1.601*
Beer	1.071	1.196	Hlth	1.740**	1.400*	ElcEq	1.590*	1.032
Smoke	2.755***	2.306***	Chems	0.659	0.544	Autos	1.444*	1.501*
Games	1.863**	1.224*	Txtls	1.325*	1.150	Carry	1.640*	1.569*
Books	1.575*	0.022	Cnstr	1.706*	1.747**	Mines	-0.087	-0.119
Hshld	1.502*	0.930	Steel	1.915**	1.396*	Coal	1.691**	1.706**
Oil	0.891	0.867	Trans	1.868**	1.275*	BusEq	2.133***	1.521*
Util	2.425***	1.4805*	Whsls	1.517*	1.978**	Paper	0.947	0.860
Telcm	1.336*	0.055	Rtail	1.340*	1.288*	Fin	1.221*	0.994
Servs	0.578	0.509	Meals	2.018**	1.671**	Other	0.815	0.022

From the above table we can see that most of the $ENC - T$ test statistics are significant, this means that our model contains useful information that is not captured by OLS settings. We can also see that MC QR model with distribution assumptions performs slightly better than the one without distribution assumptions. This could indicate that although they perform indistinguishably for in sample analysis, the distribution assumption might be useful in predictions. However, this argument needs to be further tested.

We can see that the Markov chain quantile regression setting outperforms, both in sample and out sample, OLS method. However, how to turn this advantage into construction of higher expected return portfolio still needs to be further studied.

Chapter 6

Conclusions and Possible Future Directions

In this study we explore the application of quantile regression in different areas. Quantile regression is a distribution free method, so it is applicable in the case where the distribution of the response variable exhibits some particular characteristics. The distributions of quite a few economic variables are proved to be non-normal. Quantile regression also gives us a more complete picture of the conditional distribution of the response variable, therefore enables us to examine the relationship between factors and different parts of the distribution of the variable of interest, such as tail behaviors. Therefore we try to apply quantile regression into finance study.

We develop a model that combines the market evolution dynamics with the conditional distribution of the forecasted variable. We use a three state Markov chain to model the market movement, and select different quantiles to identify the states. The distributions given different states are also modeled, and we use Baye's rule to give the mean of the marginal distribution. The prediction is based on

consistent estimates of the regression and distribution parameters. Our model is a generalization of previous work by L. Ma and L. Pohlman (2008). The new idea in our model is that we introduce the Markov chain to model the dependence of the data, while in previous work, this dynamic is assumed to be either totally dependent or independent.

We also study the theoretical properties of our model. We use AR model to capture the autocorrelation structure of time series data, and the assumption of the error terms is weakened to martingale difference sequence. With the tool of central limit theory for time series data, we show the consistency of the quantile regression parameter estimations. Furthermore, we also deduce the asymptotic properties of the estimation of Markov chain parameters. Particularly, we prove the consistency and asymptotic normality of the transition probability estimates.

Before we conduct empirical study, we also perform simulations with our model. One concern before the simulation is that we do not have a regular asymptotic distribution of the quantile regression estimates. This may cause issues when we perform statistical inference about the estimations. The simulation study aims to obtain a clearer idea of this asymptotic distribution. The results show that under some mild conditions, it is close to a normal distribution. Therefore, in application, the confidence interval estimation based on asymptotic normality gives us a satisfactory approximation.

We further apply our specification in two empirical studies. The first one applies the method in the forecast of equity risk premium. The results of individual models confirm the previous conclusion that individual predictor is unable to deliver consistent superior out of sample performance. While applying the combination methods improves the results significantly. The analysis show that our model reduces mean squared prediction error around twice as much as other competing models. The new

information included in our model is valuable in prediction and the economic evaluations also show that our model is practical. The other study examines Fama-French three factor model for both in sample relationship and out of sample prediction. The empirical results show that our model fits the data better and has greater predictive power.

Our work extends previous study on the application of quantile regression in finance by proposing a methodology that utilizes the information of market evolution process. We believe that there exist more potential applications and generalizations of our method. Many interesting questions remain open and various application are still remained to develop in this area.

One possible extension can be more sophisticated model of market evolution dynamics. Instead of three state Markov chain in our present work, we can generalize it into higher degree of dependence. Our tentative model of Markov chain has shown that introducing the market evolution provides useful information for understanding the relationship among economics variables. Therefore it is expected that the better modeling of the market evolving process may increase the degree of model fitness and prediction accuracy. Markov chain is one of the simplest forms of modeling dependence. In the current set up, the market states are only allowed to be related with consecutive time periods. However, we know that the evolution of the market is really complicated and there may exist higher degree of dependency. Therefore, we may try higher order Markov chain or apply a separate time series regression to model the market evolution process.

Once we introduce more complicated models for market evolution, other parts of our model specification may also be changed accordingly. For the model specification part, we need to change the form of probability matrix, for example, if we apply a second order Markov chain, the probability matrix will become a 9 by 9 matrix,

or come up with a combined model for the prediction, for example, if we apply a separate time series model. For the prediction part, we may apply multi-steps ahead forecast. Since we assume the degree of dependency is increased, it is reasonable to expect the information in current time period economic variables is related with future values after several time periods. For the model evaluation part, the model comparison criteria will be adjusted to higher order case.

The other direction of our future study is the quantile selection. In our present work, we select extreme quantiles for Fama French model to better study the tail behaviors and less extreme quantiles for equity risk premium forecast to keep consistent with other studies and to keep the size of sub-samples relatively moderate. Although we try different quantiles to test the robustness of the model, we do not have an algorithm to come up with an optimal set of quantiles. Usually we can use economic knowledge to define the percentage of bad and good days of the market, however, the optimization within certain range is also desirable.

Furthermore, we can also generalize our model into multi-state case. The more quantiles we estimate, the more complete picture of the distribution we obtain, however, at the cost of estimating extra parameters. At the relatively low range, the fixed weight quantile regression method performs better as the number of quantiles included in the model increases. We expect our model exhibit the same property. However, one problem we need to solve for the generalization is how to the define conditional distribution and its first and second order moments when there are more than three states. We may try some truncated distributions for example.

Finally, the Fama French project provides basis of future research in portfolio construction. We know the aim of better prediction of portfolio excess return is to adjust the portfolio locations accordingly and expect a higher return in the future. Therefore, how to turn better return prediction accuracy into higher future return

is of great importance. The widely used way of calculating the weight of each security in a portfolio involves second order estimation of the expected return. Now in our set up, we specify different distributions of excess return for different states, therefore, for different market conditions, the variance of return is assumed to be different. For each portfolio sector, we estimate the variance by dividing the whole sample into three sub-samples, each corresponding to a state. The estimation of the covariance among securities is more complicated under our model specification. We know that under bad market condition, such as financial crisis period, the covariance among different types of securities are extremely highly positive. Therefore, it is reasonable to assume different covariance matrix for different states. However, the states are estimated by comparing the excess return of the each portfolio sector with the estimated quantiles. Therefore the estimation of the market state for different portfolio might be inconsistent. How to divide the sample to estimate the covariance at different states becomes an interesting question. This is the main obstacle on our way to continue our work in the field of portfolio construction.

Bibliography

- [1] Aiolfi, M., and C. A. Favero. 2005. "Model Uncertainty, Thick Modelling and the Predictability of Stock Returns." *Journal of Forecasting* 24:233-54.
- [2] Bates, J. M., and C.W.J. Granger. "The combination of forecasts." *Operational Research Quarterly*, 20 (1969), 451-468.
- [3] Baur, D. G.; T. Dimp; and R. Jung. "Stock return autocorrelations revisited: A quantile regression approach." *Journal of Empirical Finance*, (2012), forthcoming.
- [4] Campbell, J. Y. "Stock Returns and the Term Structure." *Journal of Financial Economics*, 18 (1987), 373-99.
- [5] Campbell, J. Y., and J. H. Cochrane. "By force of habit: A consumption-based explanation of aggregate stock market behavior." *The Journal of Political Economy*, 107 (1999), 205-251.
- [6] Campbell, J. Y., and R. J. Shiller. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*, 43 (1988), 661-76.
- [7] Campbell, J. Y., and R. J. Shiller. "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*, 24 (1998), 11-26.

- [8] Campbell, J. Y., and S. B. Thompson. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?." *Review of Financial Studies*, 21 (2008), 150931.
- [9] Clark, T. E., and K. D. West. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics*, 138 (2007), 291311.
- [10] Della Corte, P.; L. Sarno; and I. Tsiakas. "An Economic Evaluation of Empirical Exchange Rate Models." *Review of Financial Studies*, 22, 9 (2009), 3491-3530.
- [11] Diebold, F. X., and R. S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13 (1995), 253.63.
- [12] Fama, E. F., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 325.
- [13] Fama, E. F., and K. R. French. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics*, 25 (1989), 23.49.
- [14] Galin L. J. "On the Markov Chain Central Limit Theorem." *Probability Surveys*, 9(2004).
- [15] Gastwirth, J.L. "On robust procedures." *Journal of the American Statistical Association*, 61(1966), 929-948.
- [16] Granger, C. W. J., and P. Newbold . "Some Comments on the Evaluation of Economic Forecasts." *Applied Economics*, 5 (1973), 35-47.
- [17] Granger, C. W. J., and R. Ramanathan. "Improved methods of combining forecasts." *Journal of Forecasting*, 3 (1984), 197-204.

- [18] Harvey, D. I.; S. J. Leybourne; and P. Newbold. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*, 16 (1998), 254-59.
- [19] Hendry, D. F., and M. P. Clements. "Pooling of Forecasts." *Econometrics Journal*, 7 (2004), 1-31.
- [20] Hsiao, C., and S. K. Wan. "Is there an optimal forecast combination?." *Journal of Econometrics*, (2012), forthcoming.
- [21] Koenker, R. "Quantile regressions." Cambridge University Press, Cambridge (2005).
- [22] Koenker, R., and G. Bassett. "Regression Quantiles." *Econometrica*, 46 (1978), 33-50.
- [23] Koenker, R., and Z. Xiao, "Unit Root Quantile Regression Inference.", *Journal of the American Statistical Association*, 99(2004), 775-787.
- [24] Koenker, R., and Z. Xiao, "Quantile Autoregression.", *Journal of the American Statistical Association*, 101(2006), 980-990.
- [25] Koenker, R., and Z. Xiao, "A Rejoinder: Quantile Autoregression.", *Journal of the American Statistical Association*, 101(2006), 1002-1006.
- [26] Kothari, S., and J. Shanken. "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." *Journal of Financial Economics*, 44 (1997), 169-203.
- [27] Lettau, M., and S. C. Ludvigson. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance*, 56 (2001), 815-49.

- [28] Ludvigson, S. C., and S. Ng. "The empirical risk-return relation: a factor analysis approach." *Journal of Financial Economics*, 83 (2007), 171222.
- [29] Ma, L., and L. Pohlman. "Return forecasts and optimal portfolio construction: a quantile regression approach." *European Journal of Finance*, 14, 5 (2008), 409425
- [30] Marquering, W., and M. Verbeek. "The Economic Value of Predicting Stock Index Returns and Volatility." *Journal of Financial and Quantitative Analysis*, 39 (2004), 40729.
- [31] Meligkotsidou, L.; E. Panopoulou; I. D. Vrontos; and S. D. Vrontos. "A Quantile Regression Approach to Equity Premium Prediction" Available online at SSRN <http://ssrn.com/abstract=2061036>, (2012)
- [32] Meligkotsidou, L.; I. D. Vrontos; and S. D. Vrontos. "Quantile Regression Analysis of Hedge Fund Strategies." *Journal of Empirical Finance*, 16 (2009), 264-279.
- [33] Menzly, L.; T. Santos; and P. Veronesi. "Understanding Predictability." *Journal of Political Economy*, 112, 1 (2004), 1-47.
- [34] Neely, C. J.; D. E. Rapach; J. Tu; and G. Zhou. "Forecasting the equity risk premium: the role of technical indicators." Federal Reserve Bank of St. Louis Working Paper 2010-008C, (2011).
- [35] OHagan, A., and J. Forster. "Bayesian Inference." In *Kendalls Advanced Theory of Statistics*, II edn, Vol. 2B, Arnold, London (2004).
- [36] Pedersen, T. Q. "Predictable return distributions." Available online at SSRN <http://ssrn.com/abstract=1658394>,(2010).

- [37] Rapach, D.; J. Strauss; and G. Zhou. "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23, 2 (2010), 821-862.
- [38] Rapach, D., and G. Zhou. "Forecasting stock returns." In preparation for G. Elliott and A. Timmermann, Eds., *Handbook of Economic Forecasting*, Vol. 2, (2011).
- [39] Spiegel, M. "Forecasting the equity premium: Where we stand today." *The Review of Financial Studies*, 21 (2008), 1453-1454.
- [40] Stock, J. H., and M. W. Watson. "Combination Forecasts of Output Growth in a Seven-Country Data Set." *Journal of Forecasting*, 23 (2004), 405-30.
- [41] Taylor, J. "A Quantile Regression approach to estimating the distribution of multiperiod returns." *Journal of Derivatives*, 7 (1999), 64-78.
- [42] Taylor, J. W. "Forecasting daily supermarket sales using exponentially weighted quantile regression." *European Journal of Operational Research*, 178 (2007), 154-167.
- [43] Timmermann, A. "Forecast combinations." In *Handbook of Economic Forecasting*, Vol. I, G. Elliott, C. W. J. Granger and A. Timmermann, eds. Amsterdam, Elsevier (2006).
- [44] Tukey, J. W. "Explanatory Data Analysis." Addison-Wesley, Reading, MA (1977).
- [45] Xiao, Z. "Quantile Cointegrating Regression", *Journal of Econometrics*, 150, 2, 6(2009), 248-260

- [46] Wang, Y. "THREE ESSAYS ON TIME SERIES QUANTILE REGRESSION.", PhD Dissertation, 7(2012)
- [47] Yu, K., and J. Zhang. "A Three-Parameter Asymmetric Laplace Distribution and Its Extension." *Communications in Statistics - Theory and Methods*, 34 (2005), 1867-1879.
- [48] Yu, K.; Z. Lu; and J. Stander. "Quantile regression: applications and current research areas." *The Statistician*, 52 (2003), 331-350.
- [49] Yu, K., and R. A. Moyeed. "Bayesian quantile regression." *Statistics and Probability Letters*, 54 (2001), 437-447.
- [50] Wachter, J., and M. Warusawitharana. "Predictable returns and asset allocation: Should a skeptical investor time the market?." *Journal of Econometrics*, 148, 2 (2009), 162-178.
- [51] Welch, I., and A. Goyal. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455-1508.
- [52] West, K. D. "Asymptotic Inference About Predictive Ability." *Econometrica*, 64 (1996), 1067-1084.
- [53] West, K.; H. Edison; and D. Cho. "A Utility-based Comparison of Some Models of Exchange Rate Volatility." *Journal of International Economics*, 35 (1993), 23-46.