A RISK RETURN RELATION IN STOCK MARKETS: EVIDENCE FROM A NEW SEMI-PARAMETRIC GARCH-IN-MEAN MODEL

A Dissertation

by

NAPON HONGSAKULVASU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ke-Li Xu
Committee Members,	Qi Li
	Dennis W. Jansen
	Ximing Wu
Head of Department,	Timothy Gronberg

August 2015

Major Subject: Economics

ABSTRACT

In this paper, I propose a new semi-parametric GARCH-in-Mean model. Since many empirical papers have the mix results on the risk-return relation, the cause of problem may come from the misspecification of conditional mean equation or conditional variance equation or both of them. My model uses non-parametric estimation in conditional mean equation and semi-parametric estimation in conditional variance equation which allows the non-linear risk return relation in conditional mean equation and allows the non-linear relation between the volatility and the cumulative sum of exponentially weighted past returns. Three parameters on my model are GARCH parameter, the leverage effect parameter and leptokurtic parameter. I also extend my model to include four exogenous variables, dividend yield, term spread, default spread and momentum into conditional mean equation by using additive model which allows each variable to have non-linear relation with the return. An empirical study on S&P 500 suggests that risk has a small affect on market return. However, when four exogenous variables are added to the model, my model shows that the risk-return relation has a positive hump shape.

DEDICATION

To my father

TABLE OF CONTENTS

1. INTRODUCTION

In the area of finance and time series econometrics, one of the most interesting topics is the relation between risk and return. Merton (1973) proposes that the relation between the expected returns and their variance is positive and linear pattern as shown in equation (1.1). Merton's intertemporal capital asset pricing model (ICAPM) explains that when investors face the additional risk, they will expect for more return in order to compensate the risk.

$$\mathbb{E}[(r_{mt} - r_{ft}) \mid \mathcal{F}_{t-1}] = \gamma \text{VAR}[(r_{mt} - r_{ft}) \mid \mathcal{F}_{t-1}] = \gamma \sigma_t^2 \tag{1.1}$$

where r_{mt} is the returns on the market portfolio and r_{ft} is the returns on risk-free asset.

However, there are many empirical papers study on this topic and surprisingly the results still unclear. Some papers report the positive relation on risk-return trade off and some papers report the negative relation, for example, the result from Backus and Gregory (1993), French et al. (1987), Gennotte and Marsh (1993), Lee et al. (2001), Lundblad (2007), Theodossiou and Lee (1995) and Whitelaw (2000). One of the most important papers in this area is Engle et al. (1987)'s paper. Engle et al. (1987) propose ARCH-in-Mean model which captures the risk-return relation by insert the conditional variance into the conditional mean equation as it's shown in equation (1.2a) and the conditional variance is determined by the previous lagged of error term from the conditional mean equation as it's shown in equation (1.2c).

$$y_t = c + \gamma\sigma_t^2 + \epsilon_t \tag{1.2a}$$

$$\epsilon_t = \xi_t\sigma_t \tag{1.2b}$$

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 \tag{1.2c}$$

where y_t is the excess returns, $\mathbb{E}(\epsilon_t \mid \mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\epsilon_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2$, σ_t^2 is the conditional variance of the excess returns and $\xi_t \sim i.i.d.(0, 1)$.

Based on ARCH-in-Mean model, the conditional variance equation (1.2c) can be easily turned to Bollerslev(1986)'s GARCH model as $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. In order to capture the leverage effect, the conditional variance equation (1.2c) can be modified to Nelson(1991)'s EGARCH model as $\log\sigma_t^2 = \omega + \alpha(\xi_{t-1} + \eta|\xi_{t-1}| + \beta\log\sigma_{t-1}^2)$ and also can be modified to Glosten et al.(1993)'s GJR-GARCH model as $\sigma_t^2 = \omega + \alpha(\epsilon_{t-1}^2 + \eta\epsilon_{t-1}^2\mathbb{1}_{(\epsilon_{t-1}<0)}) + \beta\sigma_{t-1}^2$. In both EGARCH and GJR-GARCH model, the parameter η will capture the asymmetric effects between good news and bad news.

Base on parametric ARCH-in-Mean model in equations (1.2), one of the causes of getting mixed results may come from the misspecification of the model's functional form. Since the ARCH-in-Mean has two main equations, equation (1.2a) is the conditional mean equation which captures the risk-return relation and equation (1.2c) shows the factors that determine the conditional variance. The misspecification problem may come from the conditional mean equation or the conditional variance equation or both of them.

To overcome the misspecification problem, we can apply the non-parametric estimation to the GARCH-in-Mean model. Linton and Perron (2003) propose to use the non-parametric estimator in the conditional mean equation to allow for the flexibility of functional form of risk-return relation. So they use a smooth and unknown function $\mu(\cdot)$ in equation (1.3) instead of the linear model in equation (1.2a) in order to allow the functional form of mean equation to be flexible.

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{1.3}$$

However, for the conditional variance equation (1.2c) , Linton and Perron (2003) still use the parametrically E-GARCH which has the benefit on capturing the leverage effect and still allow conditional variance to be highly persistent. In their paper, they also report that the monthly excess returns on CRSP data have a nonlinear relation with their risk. The risk-return relation appears to be hump-shape.

Conrad and Mammen (2008) propose the iterative semi-parametric approach which applies a non-parametric in the conditional mean equation. They also propose the test for parametric specification. According to their test, they report that the risk-return relation is linear on the monthly CRPS excess return data which is support the prediction of the ICAPM. However, they found the non-linear relation on the daily data which is support Linton and Perron (2003)'s results.

Christensen et al.(2012) argue that Conrad and Mammen (2008) algorithm has some questionable point. Since Conrad and Mammen (2008)'s model requires a consistent estimator for the starting values, the Quasi Maximum Likelihood (QML) estimator they use is actually inconsistent if the risk-return relation is indeed non-linear. In order to solve this problem, Christensen et al. (2012) come up with the model which does not rely on starting consistent estimator. Instead of using the traditional GARCH model, Christensen et al. (2012) apply the double autoregressive model of Ling (2004) which is using y instead of ϵ in the conditional variance equation as shown in equation (1.4).

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{1.4}$$

Zhang et al.(2013) support the idea of Ling (2004) and Christensen et al. (2012). They show that QMLE for the GARCH-in-Mean model that uses y instead of ϵ in the conditional variance equation is asymptotically normal. Their simulations and empirical results show that the estimation performs well and has comparable performance compare to the traditional GARCH-in-Mean model.

Linton and Perron (2003) state that the semi-parametric model which based on GARCH structure cannot define $\hat{\mu}(\sigma_t^2)$ in equation (1.3) so easily because σ^2 depends on lagged ϵ which in turn depends on lagged μ. Therefore, the idea of Ling (2004)'s double autoregressive model which replacing ϵ with y in equation (1.4) can overcome this difficulty.

For the literatures above, even though they use non-parametric estimation in conditional mean equation but they still leave the conditional variance equation to be parametric model. Regardless the risk return relation and set the mean equation (1.2a) as $y_t = \epsilon_t$, there are some literatures study on the non-parametric method in Engle (1982)'s ARCH and Bollerslev (1986)'s GARCH model. Pagan and Schwert (1990) and Pagan and Hong (1991) propose the nonparametric technique in ARCH model where equation (1.2c) can be set as $\sigma_t^2 = f(y_{t-1}, y_{t-2}, ..., y_{t-d})$ and $f(\cdot)$ is a smooth but unknown function. However, the information set contains a limit number of lagged which contrasts to the fact that most of the financial data have a highly persistent conditional variance.

Engle and Ng (1993) propose the PNP or partially non-parametric in GARCH model where $\sigma_t^2 = \beta\sigma_{t-1}^2 + f(y_{t-1})$ and $f(\cdot)$ is a smooth but unknown function. Linton and Mammen (2005) follow the same idea and propose a semi-parametric

ARCH (∞) model where $\sigma_t^2(\theta, f) = \sum_{j=1}^{\infty} \psi_j(\theta) f(y_{t-j})$ and $f(\cdot)$ is a smooth but unknown function. The coefficients $\psi_j(\theta) \geq 0$ and $\sum_{j=1}^{\infty} \psi_j(\theta) < \infty$. Li et al. (2005) propose to use the nonparametric series method in the conditional variance equation and they found a significant negative risk-return relation in 6 out of 12 markets. Audrino and Bühlmann (2001) and Bühlmann and McNeil (2002) propose their non-parametric GARCH model where $\sigma_t^2 = f(y_{t-1}, \sigma_{t-1}^2)$ and $f(\cdot)$ is a smooth but unknown function.

$$\sigma_t^2 = f(y_{t-1}) + \beta f(y_{t-2}) + \beta^2 f(y_{t-3}) + ... + \beta^{t-1} f(y_0) \tag{1.5a}$$

$$\sigma_t^2 = g\left\{ \sum_{j=1}^{t} \beta^{j-1} f(y_{t-j}; \eta) \right\} \tag{1.5b}$$

$$\sigma_t^2 = g(U_{t-1}) \tag{1.5c}$$

where $U_t = \sum_{j=0}^{t} \beta^j f(y_{t-j}; \eta)$

Yang (2006) proposes the semi-parametric extension of the GJR-GARCH model. The idea is σ_t^2 in equation (1.2c) can be expressed as it shown in equation (1.5a). For the GARCH(1,1) model, $f(y)$ in (1.5a) can be shown that $f(y) \equiv \alpha y^2 + \omega$. In the case of GJR-GARCH model, $f(y) \equiv \alpha(y^2 + \eta y^2 \mathbb{1}_{(y<0)}) + \omega$. Both GARCH and GJR-GARCH model can be consider as the sub-model of Yang (2006)'s semiparametric GARCH model. To allow function $g(\cdot)$ in the equation (1.5b)and (1.5c) to be flexible, Yang (2006) uses the non parametric technique in the conditional variance equation to estimate function $g(\cdot)$. The smooth and unknown function $g(\cdot)$ allows the non-linear relation between the returns volatility and the cumulative sum of exponentially weighted past returns. Yang (2006) applies his semi-parametric GARCH model to study on the foreign exchange market and found that his model outperforms the parametric GJR-GARCH model and GARCH(1,1) model.

Mishra, Su and Ullah (2010) propose a combined semi-parametric estimator, which incorporates the parametric and non-parametric estimators of the conditional variance in a multiplicative way. They show the benefit of their model over pure non-parametric model and pure parametric model. However, they use parametric estimator in conditional mean model to capture risk-return relation in the conditional mean equation.

In this paper, we propose a new semi-parametric GARCH-in-Mean model which allows the functional form in both conditional mean equation and conditional variance equation to be flexible. Since Christensen et al.(2012)'s apply non-parametric method only in the conditional mean equation and leave the conditional variance equation follows GARCH(1,1) process. We extend Christensen et al.(2012)'s model by using the non-parametric estimation in conditional mean equation and Yang (2006)'s semiparametric GARCH model in conditional variance equation. For the conditional variance equation, we follow Christensen et al.(2012) and Ling (2004) by using y instead of ϵ as it's shown in equation (1.4). The benefit of using y instead of ϵ is that the model estimation will not rely on any initial value from parametric estimation.

2. MODEL

In this paper, we extend Christensen et al.(2012)'s semiparametric GARCH-in-Mean model by using the non-parametric estimation in conditional mean equation and Yang (2006)'s semiparametric GARCH model in conditional variance equation. Christensen et al.(2012) use non-parametric estimation only in conditional mean equation and leave the conditional variance equation to follow parametrically GARCH(1,1) as it's shown below.

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{2.1a}$$

$$\epsilon_t = \xi_t \sigma_t \tag{2.1b}$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2.1c}$$

In equation (2.1c), Christensen et al.(2012) follow the idea of Ling (2004)'s Double AutoRegressive model by using y instead of ϵ. The traditional GARCH-in-Mean model has a difficulty that $\mu(\sigma_t^2)$ in equation (2.1a) cannot be estimated so easily because σ^2 depends on lagged ϵ which in turn depends on lagged $\mu(\cdot)$. By replacing ϵ with y, then σ^2 will not depend on the function $\mu(\cdot)$.

To capture the leverage effect, we follow Glosten et al.(1993)'s GJR-GARCH model by replace equation (2.1c) with

$$\sigma_t^2 = \omega + \alpha(y_{t-1}^2 + \eta y_{t-1}^2 \; \mathbb{1}_{(y_{t-1}<0)}) + \beta \sigma_{t-1}^2 \tag{2.2}$$

Then, we follow Yang (2006)'s idea by transforming equation (2.2) to

$$\sigma_t^2 = \frac{\omega}{(1-\beta)} + \alpha\left\{\sum_{j=1}^{t}\beta^{j-1}v(y_{t-j};\eta)\right\} \tag{2.3a}$$

$$\sigma_t^2 = \frac{\omega}{(1-\beta)} + \alpha\{U_{t-1}\} \tag{2.3b}$$

where $U_{t-1} = \sum_{j=1}^{t}\beta^{j-1}v(y_{t-j};\eta)$ and $v(y;\eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$.

We can see that equation (2.1c) can be transform to equation (2.3b) in which presents the linear relation between the conditional variance σ_t^2 and the cumulative sum of exponentially weighted past returns U_{t-1}. We follow Yang (2006)'s idea by relax the the functional form in equation (2.3b) as it's shown below.

$$\sigma_t^2 = g(U_{t-1}) \tag{2.4}$$

where $g(\cdot)$ is a smooth and unknown function and can be estimated by non-parametric method.

Now we can turn Christensen et al.(2012)'s semiparametric GARCH-in-Mean model to our new semiparametic GARCH-in-Mean model as it's shown below.

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{2.5a}$$

$$\epsilon_t = \xi_t \sigma_t \tag{2.5b}$$

$$\sigma_t^2 = g(U_{t-1}) \tag{2.5c}$$

where $U_{t-1} = \sum_{j=1}^{t}\beta^{j-1}v(y_{t-j};\eta)$, $v(y;\eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$ and $\xi_t \sim i.i.d.(0,1)$.

Normally, in parametric GARCH model, ϵ_t is assumed to has a normal distribution, such as Engle (1982). However some studies, such as Nelson (1991), show that

ϵ_t has a thick tail distribution. So we follow Li et al. (2005) and Linton and Perron (2003) by assume that ϵ_t has a generalized error distribution (GED). The density function of the generalized error distribution is shown here,

$$f(\epsilon) = \nu\{exp[-0.5|(\epsilon/\sigma)/\lambda|^{\nu}]\}\{\lambda 2^{(1+1/\nu)}\Gamma(1/\nu)\}^{-1} \tag{2.6}$$

where $\Gamma(\cdot)$ is the gamma function, and $\lambda \equiv [2^{(-2/\nu)}\Gamma(1/\nu)/\Gamma(3/\nu)]^{1/2}$.

For GED, ν is a tail-thickness parameter. We have a standard normal distribution when $\nu = 2$. When $\nu < 2$, the distribution has a thicker tails than the normal distribution. When $\nu > 2$, the distribution has a thinner tails than the normal distribution. We will get a double exponential distribution when $\nu = 1$ and uniformly distribution when $\nu = \infty$.

For the big picture, We need to estimate three parameters, β, η and ν, then we can calculate $U(\cdot)$. After we know $U(\cdot)$, We can estimate function $g(\cdot)$ by non-parametric method which will consequently give us the estimated conditional variance, $\hat{\sigma}_t^2$. Then, We can estimate function $\mu(\cdot)$ by non-parametric method which reveals the relationship between risk and return.

3. ESTIMATION

Our new semi-parametic GARCH-in-Mean model is presented here,

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{3.1a}$$

$$\epsilon_t = \xi_t \sigma_t \tag{3.1b}$$

$$\sigma_t^2 = g(U_{t-1}) \tag{3.1c}$$

where $U_{t-1} = \sum_{j=1}^{t} \beta^{j-1} v(y_{t-j};\eta)$, $v(y;\eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$ and $\xi_t \sim i.i.d.(0,1)$.

Suppose that we know the true value of parameters β and η. Then, we can calculate $v(y;\eta) = (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$ and $U_t(y;\beta,\eta) = \sum_{j=0}^{t} \beta^j v(y_{t-j};\eta)$. By substitute equation (3.1c) into equation (3.1a), we have $y_t = \mu(g(U_{t-1})) + \epsilon_t$. Let define the function $F(\cdot) = \mu(g(\cdot))$, then we have

$$y_t = F(U_{t-1}) + \epsilon_t \tag{3.2}$$

First, we simply estimate function $F(\cdot)$ by non-parametrically regress y_t on U_{t-1}. After we know function $\hat{F}(U)$, we can find $\hat{\epsilon}_t^2 = y_t - \hat{F}(U_{t-1})$. Since $\mathbb{E}(\epsilon_t^2 | U_{t-1} = u) = \sigma_t^2 = g(u)$, we can simply estimate function $g(\cdot)$ by non-parametrically regress $\hat{\epsilon}_t^2$ on U_{t-1}. After we know function $\hat{g}(U)$, we can easily find $\hat{\sigma}_t^2 = \hat{g}(U_{t-1})$. Then, we come back to equation (3.1a) where $y_t = \mu(\sigma_t^2) + \epsilon_t$ and non-parametrically regress y_t on $\hat{\sigma}_t^2$.

For the big picture, we have 3 smooth and unknown functions to be estimated, $\hat{F}(\cdot)$, $\hat{g}(\cdot)$ and $\hat{\mu}(\cdot)$. Normally, we can nonparametrically estimate these functions by a simple kernal estimation as it was done by Christensen et al.(2012) and Yang

(2006). However, the performance of kernal estimation is heavily depending on the bandwidth and to do the least square cross validation bandwidth selection is a very time consuming. In this paper, I decide to use the generalized additive model (GAM) of Wood (2006) to estimate the function $\hat{g}(\cdot)$ and $\hat{\mu}(\cdot)$.

To understand the GAM, let consider a simple model as $y_i = f(x_i) + \epsilon_i$. Then we assumed that function $f(\cdot)$ has a polynomial basis, so $f(x) = \sum_{j=1}^{q} b_j(x)\beta_j$. In the case of a cubic polynomial, we have $b_1(x) = 1$, $b_2(x) = x$, $b_3(x) = x^2$, $b_4(x) = x^3$, then $f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3$. Then the cubic spline is basically a connection of multiple cubic regressions and we call the connection point as a "knot". The performance of regression splines is heavily depending on the locations and the number of knots.

Wood (2006) suggests to use the penalized regression splines. The idea is we can keep the number of knots fixed, at a size a little larger than it is believed to be necessary. Then, we adding a "wiggliness" penalty to the least squares fitting objective function as

$$\|y - X\beta\| + \lambda \int_0^1 [f''(x)]^2 dx \tag{3.3}$$

In penalized regression splines, we have parameter λ as a smoothing parameter which control the tradeoff between model fit and model smoothness. When $\lambda \to \infty$, it will become a straight line estimate of $f(\cdot)$ and when $\lambda \to 0$, it will become an un-penalized regression spline estimate. Now we can see that the performance of the model depend on how we estimate the value of a smoothing parameter. If λ is too high, our model will be over smoothed. If λ is too low, our model will be under smoothed. Wood (2006) suggest that we can choose λ to minimize the generalized cross validation score.

$$GCV = \frac{n\sum_{i=1}^{n}(y_i - \hat{f}_i)^2}{[tr(I-A)]^2} \tag{3.4}$$

With the penalized regression splines, we can easily estimate $\hat{F}(\cdot)$ in equation (3.2) and $\hat{g}(\cdot)$, $\hat{\mu}(\cdot)$ in equation (3.1). One of the benefits of using Wood (2006)'s GAM is we can easily extend our model by adding more exogenous variables into the conditional mean equation. Christensen et al.(2012) put four exogenous variables linearly into the conditional mean equation as it's shown below.

$$y_t = \mu(\sigma_t^2) + \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \alpha_3 x_{3,t} + \alpha_4 x_{4,t} + \epsilon_t \tag{3.5}$$

where the four explanatory variables in Christensen et al.(2012) are the dividend yield, term spread, default spread and momentum.

By using Wood (2006)'s GAM, we can relax the linear assumption on those exogenous variables on Christensen et al.(2012). Then we will have the additive model on the conditional mean equation.

$$y_t = \mu(\sigma_t^2) + f_1(x_{1,t}) + f_2(x_{2,t}) + f_3(x_{3,t}) + f_4(x_{4,t}) + \epsilon_t \tag{3.6}$$

Now suppose we don't know the parameters β and η. Let define $\gamma = (\beta, \eta)$ and $\gamma \in \Gamma$ where $\Gamma = [\beta_1, \beta_2] \times [\eta_1, \eta_2]$ and $0 < \beta_1 < \beta_2 < 1, -\infty < \eta_1 < \eta_2 < +\infty$. So, each value of γ create the unique vector $U_{\gamma,t}$. For each vector $U_{\gamma,t}$, we will get the unique function $\hat{F}_\gamma(\cdot)$, $\hat{g}_\gamma(\cdot)$ and $\hat{\mu}_\gamma(\cdot)$. We will replace γ with any $\gamma' \in \Gamma$ and observe how the estimated $\hat{F}_{\gamma'}(\cdot)$, $\hat{g}_\gamma(\cdot)$ and $\hat{\mu}_\gamma(\cdot)$ changes.

From equation (2.5b), we assume that ϵ_t has a generalized exponential distribution (GED) with the density function in equation (2.6). For any $\gamma' = (\beta', \eta') \in \Gamma$, since each value of γ will create the unique estimated $\hat{F}_\gamma(\cdot)$, $\hat{g}_\gamma(\cdot)$ and $\hat{\mu}_\gamma(\cdot)$, we will keep replacing γ with any $\gamma' \in \Gamma$ until we find the γ' that maximize the following log likelihood function.

$$\hat{\gamma} = arg \max_{\gamma' \in \Gamma} log(\frac{\nu exp[-\frac{1}{2}|\frac{z}{\lambda^{\nu}}]|}{\lambda 2^{(1+1/\nu)}\Gamma(1/\nu)}) \tag{3.7a}$$

$$z = \frac{y - \hat{\mu}_{\gamma}(\cdot)}{\sqrt{\hat{g}_{\gamma}(\cdot)}} \tag{3.7b}$$

where $\Gamma(\cdot)$ is the gamma function, and $\lambda \equiv [2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)]^{1/2}$.

The maximization of equation (3.7) will give us the estimated parameters, β, η and ν. The parameter β will tell us how the effect of past returns on volatility decays overtime. The parameter η will tell us about the leverage effect or the asymmetric effect of good news and bad news. The parameter ν will tell us about the thickness of a distribution tails.

Our semi-parametric GARCH-in-Mean model offer some advantages over Christensen et al.(2012)'s semiparametric GARCH-in-Mean model. First, we nest GJR-GARCH model into our model. Our model accounts for the leverage effect via the parameter η. Second, we relax the functional form of both conditional mean equation and conditional variance equation to be flexible while Christensen et al.(2012)allow only conditional mean equation to be relax and still keep conditional variance equation to follow parametrically GARCH(1,1) process. Third, our model allow for thick tail distribution by using generalized error distribution (GED).

3.1 Estimation Procedure

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{3.8a}$$

$$\epsilon_t = \xi_t \sigma_t \quad , \quad \epsilon_t | \Omega_{t-1} \sim GED(0, \sigma_t^2, \nu) \tag{3.8b}$$

$$\sigma_t^2 = g(U_{t-1}) \tag{3.8c}$$

where $U_{t-1} = \sum_{j=1}^{t} \beta^{j-1} v(y_{t-j};\eta)$, $v(y;\eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$ and $\xi_t \sim i.i.d.(0,1)$.

- Step:1 Estimate parameter $\hat{\gamma} = (\hat{\beta}, \hat{\eta})$ by performing equation (3.7). Then, we calculate $\hat{v}(y;\hat{\eta}) = (y^2 + \hat{\eta} y^2 \mathbb{1}_{(y<0)})$ and $\hat{U}_t = \sum_{j=0}^{t} \hat{\beta}^j \hat{v}(y_{t-j};\hat{\eta})$.
- Step:2 Estimate $F(\cdot)$ by nonparametrically regress y_t on $\hat{U}_{t-1}$.
- Step:3 After we know $\hat{F}(\cdot)$, we can calculate $\hat{\epsilon}_t^2 = y_t - \hat{F}(\hat{U}_{t-1})$. Then, we can estimate $g(\cdot)$ by nonparametrically regress $\hat{\epsilon}_t^2$ on $\hat{U}_{t-1}$.
- Step:4 After we know $\hat{g}(\cdot)$, we can calculate $\hat{\sigma}_t^2 = \hat{g}(\hat{U}_{t-1})$. Then, we can estimate $\mu(\cdot)$ by nonparametrically regress y_t on $\hat{\sigma}_t^2$.

4. SPECIFICATION TEST

Since this paper uses the generalized additive model (GAM) of Wood (2006) to estimate the model which is based on series regression, we bring two consistent specification tests to test the relationship between risk and return. We use the two specification tests from Hong and White (1995) and Sun and Li (2006). These two specification tests use nonparametric series regression as same as our model. The purpose of specification test is to test whether the risk return relation is linear or not. Let $\mathbb{E}(y|\sigma^2) = \mu_0(\sigma^2)$ and let the parametric regression model is $f(\sigma^2, \gamma)$. For the null hypothesis, the risk return relation has a linear relationship. For the alternative hypothesis, the risk return relation is not a linear pattern.

$$H_0 \qquad P(\mu_0(\sigma^2) = f(\sigma^2, \gamma_0)) = 1 \quad for \quad some \quad \gamma_0 \in \Gamma \tag{4.1a}$$

$$H_a \qquad P(\mu_0(\sigma^2) \neq f(\sigma^2, \gamma_0)) = 1 \quad for \quad some \quad \gamma_0 \in \Gamma \tag{4.1b}$$

Under H_0, Hong and White (1995) specification test is shown here

$$HW_n = \frac{(n\hat{m}_n - \hat{R}_n)}{\hat{S}_n} \xrightarrow{d} N(0,1) \tag{4.2a}$$

$$\hat{m}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{v}_t \hat{\epsilon}_t \tag{4.2b}$$

$$\hat{\epsilon}_t = y_t - f(\hat{\sigma}_t^2, \hat{\gamma}) \tag{4.2c}$$

$$\hat{v}_t = \hat{\mu}(\hat{\sigma}_t^2) - f(\hat{\sigma}_t^2, \hat{\gamma}) \tag{4.2d}$$

$$\hat{R}_n = \sum_{t=1}^{n} p^k(\hat{\sigma}_t^2)'(P'P)^{-1} p^k(\hat{\sigma}_t^2)\hat{\epsilon}_t^2 \tag{4.2e}$$

$$\hat{S}_n^2 = 2\sum_{t=1}^{n}\sum_{s=1}^{n} \{p^k(\hat{\sigma}_t^2)(PP)^{-1} p^k(\hat{\sigma}_t^2)\}^2 \hat{\epsilon}_t^2 \hat{\epsilon}_s^2 \tag{4.2f}$$

where $p^k(\hat{\sigma}_t^2)$is a $k \times 1$ vector of base functions evaluated at x_t. P is $n \times k$ with i-th row given by $p^k(\hat{\sigma}_t^{2\prime})$.

Hong and White (1995) specification test is based on the sample covariance between the parametric model's residual and the discrepancy between parametric and non-parametric estimator. Hong and White (1995) use series regression to estimate the non-parametric model. Under correct specification, Hong and White (1995) test converges in distribution to a unit normal. Under misspecification, Hong and White (1995) test diverges to infinity faster than parametric rate, $n^{-1/2}$.

Under H_0, Sun and Li (1995) specification test is shown here

$$SL_n = n\frac{\hat{I}_n}{\hat{S}_n} \xrightarrow{d} N(0,1) \tag{4.3a}$$

$$\hat{I}_n = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n} \hat{\epsilon}_t p^k(\hat{\sigma}_t^2)'(P'P)^{-1}p^k(\hat{\sigma}_s^2)\hat{\epsilon}_s \tag{4.3b}$$

$$\hat{S}_n^2 = 2\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n} \{p^k(\hat{\sigma}_t^2)(PP)^{-1}p^k(\hat{\sigma}_t^2)\}^2\hat{\epsilon}_t^2\hat{\epsilon}_s^2 \tag{4.3c}$$

$$\hat{\epsilon}_t = y_t - f(\hat{\sigma}_t^2, \hat{\gamma}) \tag{4.3d}$$

where $p^k(\hat{\sigma}_t^2)$is a $k \times 1$ vector of base functions evaluated at x_t. P is $n \times k$ with i-th row given by $p^k(\hat{\sigma}_t^{2\prime})$.

Sun and Li (1995) argue that Hong and White (1995) specification test can have finite sample bias from involving some non-zero center terms. Moreover, Hong and White (1995) test is based on under-smoothing conditions on series estimators. Then Sun and Li (1995) propose the alternative specification test which does nit have a non-zero center term and allows for optimal smoothing under general condition. Sun and Li (1995) show that their test is better than Hong and White (1995) test.

In this paper, we employ both Hong and White (1995) and Sun and Li (1995) specification test to test the relationship between risk and return. We will compare the performance of these tests in our Monte Carlo simulation and show which test is better in to test the relationship between risk and return.

5. MONTE CARLO SIMULATION

Since our model allows the functional form in both conditional mean equation and conditional variance equation to be flexible, the conditional mean equation can reveal both linear and non-linear relation between risk and return. Moreover, the conditional variance equation in our model not only nests both GARCH(1,1) and GJR-GARCH but also allow the relationship between the volatility σ_t^2 and the cumulative sum of exponentially weighted past returns U_{t-1} to be non-linear. Since our model is the extension of Christensen et al.(2012), we construct the Monte Carlo Simulation with the cases from Christensen et al.(2012) to show that even though we use non-parametric technique in both conditional mean equation and conditional variance equation, it still works very well.

$$y_t = \mu(\sigma_t^2) + \epsilon_t \tag{5.1a}$$

$$\epsilon_t = \xi_t \sigma_t \quad , \quad \epsilon_t | \Omega_{t-1} \sim GED(0, \sigma_t^2, \nu) \tag{5.1b}$$

$$\sigma_t^2 = g(U_{t-1}) \tag{5.1c}$$

where $U_{t-1} = \sum_{j=1}^{t} \beta^{j-1} v(y_{t-j}; \eta)$, $v(y; \eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$ and $\xi_t \sim i.i.d.(0, 1)$.

In our simulation, we have 6 cases, three cases for the linear risk-return relation and other three cases for the non-linear risk-return relation. All of 6 cases are from Christensen et al.(2012). For the conditional mean equation, L1-L3 are the case that risk and return has linear relation while N1-N3 are the non-linear cases.

•N1: $\mu(\sigma_t^2) = \sigma_t^2 + 0.5\sin(10\sigma_t^2)$

•N2: $\mu(\sigma_t^2) = 0.5\sigma_t^2 + 0.1\sin(0.5 + 20\sigma_t^2)$

•N3: $\mu(\sigma_t^2) = \sigma_t^2 + 0.12\sin(3 + 30\sigma_t^2)$

•L1: $\mu(\sigma_t^2) = 0.05\sigma_t^2$

•L2: $\mu(\sigma_t^2) = 0.5\sigma_t^2$

•L3: $\mu(\sigma_t^2) = \sigma_t^2$

In Christensen et al.(2012), the conditional variance follows GARCH(1,1) process but in this paper we extend it to GJR-GARCH model as $\sigma_t^2 = \omega + \alpha(\epsilon_{t-1}^2 + \eta\epsilon_{t-1}^2 \mathbb{1}_{(\epsilon_{t-1}<0)}) + \beta\sigma_{t-1}^2$. As we show before, the traditional parametric GJR-GARCH model is actually the linear relation between the conditional variance σ_t^2 and the cumulative sum of exponentially weighted past returns U_{t-1}.

$$\sigma_t^2 = \frac{\omega}{(1-\beta)} + \alpha\{U_{t-1}\} \tag{5.2}$$

where $U_{t-1} = \sum_{j=1}^{t} \beta^{j-1} v(y_{t-j};\eta)$ and $v(y;\eta) \equiv (y^2 + \eta y^2 \mathbb{1}_{(y<0)})$.

For the conditional variance equation, all 6 cases are generate from GJR-GARCH model with is the linear relation between the conditional variance σ_t^2 and the cumulative sum of exponentially weighted past returns U_{t-1} as it's shown in equation (5.2). For all 6 cases, we set $\omega = 0.01$ and $\alpha = 0.1$. For the GARCH parameter β and leverage effect parameter η, we set them as follow.

•N1: $\beta = 0.70 \quad \eta = 0.80$

•N2: $\beta = 0.87 \quad \eta = -0.80$

•N3: $\beta = 0.80 \quad \eta = 0$

•L1: $\beta = 0.90 \quad \eta = -0.80$

•L2: $\beta = 0.82 \quad \eta = 0.80$

•L3: $\beta = 0.80 \quad \eta = 0$

From equation (5.1b), we assume that $\xi_t \sim N(0,1)$ then we expect to see $\nu = 2$. For the sample size, we generate the data at $T = 3,400$ and throw the first 400 observations away in order to avoid the start-out effect, so our sample size is equal to 3,000. We repeat the simulation for 1,000 times.

For the parameters estimation, we need to estimate 3 parameters, GARCH parameter β, leverage effect parameter η and leptokurtic parameter ν. We use grid search to find the optimal parameter that maximize the GED log likelihood function. For all cases, we generate data under normal distribution assumption then the true value of $\nu_0 = 2$. For parameter ν, we set the lower bound and upper bound of grid search to be between 1.8 and 2.2 with 0.05 increment. Since β and η are set different in each case, the lower bound and upper bound of grid search will be set differently.

In case N1, $\beta_0 = 0.70$ then the lower bound and upper bound of β grid search is between 0.60 and 0.80 with 0.01 increment. We set $\eta_0 = 0.80$ then the lower bound and upper bound of η grid search is between 0 and 2 with 0.1 increment. In case N2, $\beta_0 = 0.87$ then the lower bound and upper bound of β grid search is between 0.80 and 0.99 with 0.01 increment. We set $\eta_0 = -0.80$ then the lower bound and upper bound of η grid search is between -2 and 0 with 0.1 increment. In case N3, $\beta_0 = 0.80$ then the lower bound and upper bound of β grid search is between 0.70

and 0.90 with 0.01 increment. We set $\eta_0 = 0$ then the lower bound and upper bound of η grid search is between -1 and 1 with 0.1 increment.

In case L1, $\beta_0 = 0.90$ then the lower bound and upper bound of β grid search is between 0.80 and 0.99 with 0.01 increment. We set $\eta_0 = -0.80$ then the lower bound and upper bound of η grid search is between -2 and 0 with 0.1 increment. In case L2, $\beta_0 = 0.82$ then the lower bound and upper bound of β grid search is between 0.70 and 0.90 with 0.01 increment. We set $\eta_0 = 0.80$ then the lower bound and upper bound of η grid search is between 0 and 2 with 0.1 increment. In case L3, $\beta_0 = 0.80$ then the lower bound and upper bound of β grid search is between 0.70 and 0.90 with 0.01 increment. We set $\eta_0 = 0$ then the lower bound and upper bound of η grid search is between -1 and 1 with 0.1 increment.

5.1 Monte Carlo Simulation Results

The first panel in figure A.1 to A.6 shows the simulation results in conditional mean equation from 6 cases. The blue line is the true mean function while the orange line is the pointwise median of our model. The two red lines are the pointwise 5% and 95% quantiles of our model. For the non-linear risk return relation in case N1 to N3, we can see that our estimator can reveal the true function very well as the orange line follow the blue line very closely. For the linear risk return relation in case L1 to L3, our model works very well as we expected. The second panel figure A.1 to A.6 shows a histogram of estimated conditional variance, $\hat{\sigma^2}$. We can see that when $\hat{\sigma^2}$ is dense, the red confidence band is very narrow. In contrast, when $\hat{\sigma^2}$ is sparse, the red confidence band is become wide.

For figure A.7 to A.12, we show the simulation result in conditional variance equation from 6 cases. Since all 6 cases follow GJR-GARCH process, the true function of conditional variance equation is linear as is shown in blue line. The orange line is the pointwise median of our model while the two red lines are the pointwise 5% and 95% quantiles of our model. For all 6 cases, our model work very well in order to reveal the true function of conditional variance. The second panel figure A.7 to A.12 shows a histogram of the cumulative sum of exponentially weighted past returns, U_t. We can see that when U_t is dense the red confidence band is very narrow. In contrast, when U_t is sparse the red confidence band is become wide.

In order to measure the goodness of fit, we provide the mean square error on both conditional mean equation and conditional variance equation. We calculate the mean square error by the following equation.

$$MSE(mean) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\mu}(\hat{\sigma}_t^2))^2 \tag{5.3a}$$

$$MSE(var) = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t^2 - \hat{g}(\hat{U}_{t-1}))^2 \tag{5.3b}$$

For table B.1, we show how our model's fit improve when the sample size is bigger for the case N1. As we can see, for the mean square error of conditional mean equation, when sample size increase from 500 to 3,000, our model has the same level of performance. For conditional variance equation, when the sample is increasing, the mean square error is decreasing. It means that the bigger sample will give a better result.

For the parameters estimation, our model has 3 parameters to be estimated, GARCH parameter β, leverage effect parameter η and leptokurtic parameter ν. Figure A.13 to A.18 show the results on GARCH parameter estimation for all 6 cases. The curve on these figures are the density of $\hat{\beta}$ over 1,000 times Monte Carlo simulation. The green vertical line show the true value of β_0 in each cases. For all 6 cases, we can see that the peak of density curve is located near the true value of β which means our model can estimate GARCH parameter β very well.

Figure A.20 to A.25 show the results on leverage effect parameter estimation for all 6 cases. For all 6 cases, we can see that the peak of density curve is located near the true value of η or the green vertical line which means our model can estimate leverage effect parameter η very well.

Figure A.26 to A.31 show the results on leptokurtic parameter estimation for all 6 cases. Since all 6 cases are generate on under normality assumption, we expect to see $\hat{\nu} = 2$. We can see that the peak of density curve is located near the true value of ν or the green vertical line which means our model can estimate leptokurtic parameter ν very well.

Table B.2 shows the parameters estimation result from 1,000 times Monte Carlo simulation for all 6 cases. We report the median, 5% quantiles and 95% quantiles of 3 parameters, β, η and ν. For GARCH parameter β, we can see that the median of our estimations are very close to the true values for all 6 cases. For leverage effect parameter η, the median of our estimations also perform very well for all 6 cases. For leptokurtic parameter, we expect to see $\hat{\nu} = 2$ because our data are generated from normal distribution. We can see that the median of $\hat{\nu}$ in all 3 cases are equal to 2 which mean that our model work very well on revealing the leptokurtic parameter. Moreover table B.3 shows mean and standard deviation of all 3 parameter estimation for all 6 cases.

Now we show how the parameter estimation performance responses on the sample size. We show only on case N1 with difference sample size from 500 to 3,000. From table B.4, we can see the median of our estimator is very close to the true value in all data sets. However, when the sample is small as 500, the confidence bands are wide compare to the sample size 3,000. When the sample size increase, the confidence band becomes narrow which means our model becomes more precise.

Now we bring 2 nonparametric specification tests to test the conditional mean equation on all 6 cases. We apply Hong and White (1995) and Sun and Li (2006) specification test. The reason that we use these 2 tests is that their tests are based on nonparametric series estimation as same as our model. Table B.4 shows the results of Hong and White (1995) and Sun and Li (2006) specification test on our 6 simulation cases. The numbers on Table B.5 are the percentage of rejection rate at 5%. On these 2 tests, the null hypothesis is the true function which show the relationship between risk and return is the linear pattern. For case N1 to N3, the conditional mean equation has a nonlinear relation between risk and return then we expect to see a high rejection rate of null hypothesis on these 2 tests. For case L1 to L3, the true function of the conditional mean equation is linear then we expect to see a that we fail to reject the null hypothesis.

Case L1 to L3 on table B.5 show that Hong and White (1995) and Sun and Li (2006) specification test perform very well from $T = 500$ to $T = 3,000$. The rejection rate on case L1 to L3 is very close to 5%. For case N1 to N3 on table B.5, both test perform not quite well when the sample size is small at $T = 500$. However when sample increase from $T = 500$ to $T = 3,000$, these 2 test perform significantly better. We have to note our 2 test perform quite poor on case N2 because if we see the figure A.2, most of the risk is in the range of 0.10 to 0.20 in which the true function in those area is a negative linear shape.

6. EMPIRICAL STUDY

In this section, we show the results on the real financial data. We apply our new semi-parametric GARCH-in-Mean model to Standard and Poor (S&P) 500 stock market index. For the data set, we use the daily data of S&P 500 index from January 2, 1990 to December 31, 2014 which $T = 6,246$. The data set is provided by CRSP. From our model in equation (2.5), y_t is defined as the valued weighted return include dividend.

In this section, we will have 2 parts. First, we will focus only on the risk return relation which is based on equation (2.5). In conditional mean equation, we will estimate only function $\mu(\cdot)$ to reveal the relationship between risk and return. In second, we will follow Christensen et al. (2012) by adding 4 exogenous variables, dividend yield, term spread, default spread and momentum. In Christensen et al. (2012), these 4 variables are adding to their model in a linear way. However our model offer more flexibility by using Wood (2006)'s additive model which is based on nonparametric series estimation as it's shown in equation (3.6).

For the dividend yield, some paper uses it for determining the return such as Campbell and Shiller (1988a,b), Fama and French (1988,1989) and Christensen et al. (2012). We follow Christensen et al. (2012) by calculate the dividend yield from $\sum_{j>0} d_{t-j}$ over 12 months period which include $t-1$ and then divided by p_{t-1}. Campbell (1987), Fama and French (1988,1989) and Christensen et al. (2012) use the term spread to determine the return. We calculate the term spread by the difference between 10 years and 1 year treasury constant maturity rate. The default spread has been used by Fama and French (1988,1989) and Christensen et al. (2012) to determine the return. We calculate the default spread by the difference between Baa and Aaa Moody's seasoned corporate bond yield. The momentum has been used by Keim and Stambaugh (1986), Carhart (1997) and Christensen et al. (2012) to determine the return. We calculate the momentum by $logP_{t-1} - log\overline{P}$ and $\overline{P}$ is the average of the market index over 12 months which ending on period $t-1$.

For the confidence bands, we follow Linton and Perron (2003) and Christensen et al. (2012) by using the wild bootstrap. We compute the nonparametric confidence bands by the following algorithm.

- Step 1: With $\hat{\beta}, \hat{\eta}, \hat{\nu}, \hat{\mu}(\cdot), \hat{g}(\cdot)$, $\quad \hat{U}_{t-1} = \sum_{j=1}^{t} \hat{\beta}^{j-1}(y_{t-1}^2 + \hat{\eta} y_{t-1}^2 \mathbb{1}_{(y_{t-1}<0)})$,
  $\hat{\sigma}_t^2 = \hat{g}(\hat{U}_{t-1})$, $\quad \hat{\xi}_t = \frac{y_t - \hat{\mu}(\hat{\sigma}_t^2)}{\hat{\sigma}_t}$, $\quad \xi_t^c = \hat{\xi}_t - \frac{1}{T}\sum_{t=1}^{T} \hat{\xi}_t$

- Step 2: u_t, a discrete variable taking the value -1 and 1 with equal prob (0.5).
  Draw $(u_1, u_2, ..., u_T)$ $\quad , \quad \xi_t^* = \xi_t^c u_t$

- Step 3: Given initial starting value for y_0.,
  $\hat{U}_{t-1}^* = \sum_{j=1}^{t} \hat{\beta}^{j-1}(y_{t-1}^{*2} + \hat{\eta} y_{t-1}^{*2} \mathbb{1}_{(y_{t-1}<0)})$, $\quad \hat{\sigma}_t^{*2} = \hat{g}(\hat{U}_{t-1}^*)$,
  $\hat{\epsilon}_t^* = \xi_t^* \hat{\sigma}_t^*$, $\quad y_t^* = \hat{\mu}(\hat{\sigma}_t^{*2}) + \hat{\epsilon}_t^*$

- Step 4: With the bootstrapped sequence $\{y_t^*\}_{t=1}^{T}$, calculated $\hat{\beta}^*, \hat{\eta}^*, \hat{\nu}^*, \hat{\mu}(\cdot)^*, \hat{g}(\cdot)^*$ by our new semi-parametric GARCH-in-Mean model.

- Step 5: We repeat step 2 to 4 n times. The pointwise p 100% confidence band around $\hat{\mu}(\cdot)$ is calculated by $p/2$ and $(1-p)/2$ quantiles of the empirical distribution of the n bootstrapped estimates $\hat{\mu}(\cdot)^*$ of $\hat{\mu}(\cdot)$. The standard errors of $\hat{\beta}, \hat{\eta}$ and $\hat{\nu}$ are estimated from the sample standard deviation of the n bootstrapped estimates $\hat{\beta}^*, \hat{\eta}^*$ and $\hat{\nu}^*$.

6.1 Empirical Results on S&P 500

In this section, we show the result on the risk return relation which is based on equation (2.5). We apply our new semiparametric GARCH-in-Mean model to daily data of S&P 500 return. The data set of S&P 500 is shown in figure A.32. Figure A.33 shows the estimation results on the conditional mean equation on equation (2.5a). The blue line is our new semiparametric estimation while the 2 red lines are 5% and 95% quantiles curves. The second panel of Figure A.33 shows the histogram of estimated conditional variance, $\hat{\sigma^2}$. For the S&P 500 index, we find the positive and significant relation between risk and return. However, the slope of blue line is quite flat, so we can interpret that risk has a little effect on return.

Figure A.34 shows the comparison on the conditional mean equation from different model. We perform the traditional parametric GRACH-in-Mean (GM) and GJR GRACH-in-Mean (GJR) model as it's shown on brown and black line. Our new semi-parametric GARCH-in-Mean model is shown in blue line or NP-both. The yellow line (LM) is the estimation of our model in equation (2.5) but use the linear regression instead of nonparametric series regression on both function $\mu(cdot)$ in conditional mean equation and function $g(\cdot)$ in conditional variance equation. So the yellow line or LM is basically the parametric GJR GARCH-in-Mean model but using y_t instead ϵ_t in conditional variance equation as in Ling (2004). The red line (NP-mean) is the estimation of our model in equation (2.5) but use the nonparametric estimation only in conditional mean equation and leave the conditional variance equation to be estimated by linear regression. The red line or NP-mean follows Christensen et al. (2012) approach. However the difference between NP-mean and Christensen et al. (2012) is that we use the series estimator while Christensen et al. (2012) use kernel method. The reason that we use series estimator is because we want to compare their approach to our model which also based on series estimation method.

From Figure A.34, we can see that our estimator, LM and NP-mean are very close to each other while GM and GJR have a steeper slope. For GM, it's reveal the positive and significant risk return relation. However, GJR shows the positive but insignificant risk return relation. All 5 approaches confirm Merton (1973)'s intertemporal capital asset pricing model (ICAPM) that risk and return has a positive relation.

From Figure A.35 shows the result on conditional variance equation (2.5a). The blue line is our new semiparametric estimation while the 2 red lines is 5% and 95% quantiles curves. Our result shows that relation between the estimated variance $\hat{\sigma_t^2}$ and the cumulative sum of exponentially weighted past returns U_{t-1} is linear.

Table B.6 show the estimation results on 3 parameters, β, η and ν. The standard error is calculated by standard deviation of bootstrap parameter and point estimation. For the GARCH parameter, β, we get 0.92 with standard error 0.0125 which is very significant. For leverage effect parameter, η, we get 100 with very high standard error 98.99. So the leverage effect seems to be in significant for this data set. For leptokurtic parameter, ν, we get 1.4 with standard error 0.0186 which is very significant. So we have the evidence that the distribution of ϵ has a fat tail shape.

In order to compare the performance of various approaches, we follow Christensen et al. (2012) by using the following goodness of fit measures.

$$MSE(mean) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\mu}(\hat{\sigma}_t^2))^2 \tag{6.1a}$$

$$MSE(var) = \frac{1}{T}\sum_{t=1}^{T}\{(y_t - \hat{\mu}(\hat{\sigma}_t^2))^2 - \hat{\sigma}_t^2\}^2 \tag{6.1b}$$

The results on the goodness of fit are shown in Table B.7. Table B.7 shows that LM, NP-mean and our model or NP-both have the same level of fit and perform better than traditional parametric GARCH model for the conditional mean equation. For the fit on conditional variance equation, our model which relaxes functional form on both conditional mean and variance equation performs the best among 5 approaches.

6.2 Empirical Results on S&P 500 with 4 Exogeneous Variables

In these section, we follow Christensen et al. (2012) by adding 4 exogenous variables into our new semiparametric GARCH-in-Mean model. All 4 variables are assumed to be the determinate of the return then these variables are adding to conditional mean equation. We improve Christensen et al. (2012) approach by relax the function form of these 4 exogenous variables. We apply Wood (2006)'s additive model as it's shown in equation (3.6). The 4 exogenous variables are dividend yield, term spread, default spread and momentum. The data sets of these 4 variables are shown in figure A.36 to A.39.

Figure A.40 shows the estimation results on the conditional mean equation on equation (2.5a). The blue line is our new semiparametric estimation while the 2 red lines is 5% and 95% quantiles curves. The second panel of Figure A.40 shows the histogram of estimated conditional variance, $\hat{\sigma^2}$. For the S&P 500 index, we find the positive and significant relation between risk and return when risk is greater than 1.

Figure A.41 shows the comparison on the conditional mean equation from different model. Our new semi-parametric GARCH-in-Mean model is shown in blue line. The yellow line (LM) is the estimation of our model in equation (2.5) but use the linear regression instead of nonparametric series regression on both function $\mu(\cdot)$ in conditional mean equation and function $g(\cdot)$ in conditional variance equation. So the yellow line or LM is basically the parametric GJR GARCH-in-Mean model but using y_t instead ϵ_t in conditional variance equation as in Ling (2004). The red line (NP-mean) is the estimation of our model in equation (2.5) but use the non-parametric estimation only in conditional mean equation and leave the conditional variance equation to be estimated by linear regression. The red line or NP-mean follows Christensen et al. (2012) approach. However the difference between NP-mean and Christensen et al. (2012) is that we use the series estimator while Christensen et al. (2012) uses kernel method. The reason that we use series estimator is because we want to compare their approach to our model which also based on series estimation method. From Figure A.41, we can see that All 5 approaches confirm Merton (1973)'s intertemporal capital asset pricing model (ICAPM) that risk and return has a positive relation. However, we can see that our estimator reveal that the risk-return relation is not in a linear shape. Our result supports Linton and Perron (2003)'s study that risk and return relation has a hump shape.

For first panel in Figure A.42 to A.45, the blue line is our estimator on each exogenous variables while the 2 red lines are 5% and 95% quantiles curves. The second panels in A.42 to A.45 show the histogram of each exogenous variables. Figure A.42 shows the positive but insignificant relationship between dividend yield and return. Figure A.43 shows that term spread and return relation has a convex shape. When term spread is in the range between -0.5 to 1.5, term spread has a negative and significantly effect on return. When term spread is in the range between 1.5 to 3.5, term spread has a positive and significantly effect on return. Figure A.44 shows the negative and significantly relationship between default spread and return when default spread is in the range of 0.6 to 1.0. Figure A.45 shows that momentum and return relation has a hump shape.

From Figure A.46 shows the result on conditional variance equation (2.5a). The blue line is our new semiparametric estimation while the 2 red lines is 5% and 95% quantiles curves. Our result shows that relation between the estimated variance $\hat{\sigma_t^2}$ and the cumulative sum of exponentially weighted past returns U_{t-1} is linear.

Table B.8 show the estimation results on 3 parameters, β, η and ν. The standard error is calculated by standard deviation of bootstrap parameter and point estimation. For the GARCH parameter, β, we get 0.92 with standard error 0.0095 which is very significant. For leverage effect parameter, η, we get 100 with very high standard error 95.62. So the leverage effect seems to be in significant for this data set. For leptokurtic parameter, ν, we get 1.4 with standard error 0.0246 which is very significant. So we have the evidence that the distribution of ϵ has a fat tail shape.

The results on the goodness of fit are shown in Table B.9 Table B.9 shows that our model or NP-both which relaxes functional form on both conditional mean and variance equation performs the best among 5 approaches for the conditional mean equation. For the fit on conditional variance equation, our model also performs better than other 4 approaches. So the goodness of fit supports that our model which uses the nonparametric estimator on both conditional mean and variance equation is better than fix the function form on either or both conditional mean and variance equation to be a linear.

7. CONCLUSION

We propose the new semi-parametric GARCH-in-Mean model which relaxes the functional form on both conditional mean and variance equation. We use the non-parametric series estimator to capture the non-linear relation between risk and return. We extend model by using additive model to capture the nonlinear relation on 4 exogenous variables, dividend yield, term spread, default spread and momentum.

Our Monte Carlo simulation shows how our model works. We show that our model can reveal the true function on conditional mean equation whether it is linear or nonlinear. Our model can estimate the parameters quite precisely. Three parameters on our model can capture the GARCH parameter, leverage effect and leptokurtic of the distribution.

For the empirical study on S&P 500 index, we show that risk has a very small impact on market return. However, when we add 4 exogenous variables, dividend yield, term spread, default spread and momentum, our model reveals that risk and return has a positive hump shape relation which is support Linton and Perron (2003)'s study. Our model also reveals that there is no leverage effect on the market and also found the evidence that the distribution has a fat tail rather than normal.

REFERENCES

Audrino, F. and P. Bühlmann (2001) "Tree-structured generalized autoregressive conditional heteroscedastic models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 4, pp. 727–744.

Backus, D. K. and A. W. Gregory (1993) "Theoretical relations between risk premiums and conditional variances," *Journal of Business & Economic Statistics*, Vol. 11, No. 2, pp. 177–185.

Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, Vol. 31, No. 3, pp. 307–327.

Bühlmann, P. and A. J. McNeil (2002) "An algorithm for nonparametric GARCH modelling," *Computational Statistics & Data Analysis*, Vol. 40, No. 4, pp. 665–683.

Campbell, J. Y. (1987) "Stock returns and the term structure," *Journal of Financial Economics*, Vol. 18, No. 2, pp. 373–399.

Campbell, J. Y. and R. J. Shiller (1988a) "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, Vol. 1, No. 3, pp. 195–228.

——— (1988b) "Stock prices, earnings, and expected dividends," *The Journal of Finance*, Vol. 43, No. 3, pp. 661–676.

Carhart, M. M. (1997) "On persistence in mutual fund performance," *The Journal of Finance*, Vol. 52, No. 1, pp. 57–82.

Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2012) "Semiparametric inference in a GARCH-in-mean model," *Journal of Econometrics*, Vol. 167, No. 2, pp. 458–472.

Conrad, C. and E. Mammen (2008) "Nonparametric regression on latent covariates with an application to semiparametric GARCH-in-mean models," *University of Heidelberg, Department of Economics, Discussion Paper*, No. 473.

Engle, R. F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987) "Estimating time varying risk premia in the term structure: The ARCH-M model," *Econometrica: Journal of the Econometric Society*, pp. 391–407.

Engle, R. F. and V. K. Ng (1993) "Measuring and testing the impact of news on volatility," *The Journal of Finance*, Vol. 48, No. 5, pp. 1749–1778.

Fama, E. F. and K. R. French (1988) "Dividend yields and expected stock returns," *Journal of Financial Economics*, Vol. 22, No. 1, pp. 3–25.

——— (1989) "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, Vol. 25, No. 1, pp. 23–49.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987) "Expected stock returns and volatility," *Journal of Financial Economics*, Vol. 19, No. 1, pp. 3–29.

Gennotte, G. and T. A. Marsh (1993) "Variations in economic uncertainty and risk premiums on capital assets," *European Economic Review*, Vol. 37, No. 5, pp. 1021–1041.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993) "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The Journal of Finance*, Vol. 48, No. 5, pp. 1779–1801.

Hong, Y. and H. White (1995) "Consistent specification testing via nonparametric series regression," *Econometrica: Journal of the Econometric Society*, pp. 1133–1159.

Keim, D. B. and R. F. Stambaugh (1986) "Predicting returns in the stock and bond markets," *Journal of Financial Economics*, Vol. 17, No. 2, pp. 357–390.

Lee, C. F., G.-m. Chen, and O. M. Rui (2001) "Stock returns and volatility on China's stock markets," *Journal of Financial Research*, Vol. 24, No. 4, pp. 523–543.

Li, Q., J. Yang, C. Hsiao, and Y.-J. Chang (2005) "The relationship between stock returns and volatility in international stock markets," *Journal of Empirical Finance*, Vol. 12, No. 5, pp. 650–665.

Ling, S. (2004) "Estimation and testing stationarity for double-autoregressive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 66, No. 1, pp. 63–78.

Linton, O. and E. Mammen (2005) "Estimating Semiparametric ARCH (∞) Models by Kernel Smoothing Methods1," *Econometrica*, Vol. 73, No. 3, pp. 771–836.

Linton, O. and B. Perron (2003) "The shape of the risk premium: evidence from a semiparametric generalized autoregressive conditional heteroscedasticity model," *Journal of Business & Economic Statistics*, Vol. 21, No. 3, pp. 354–367.

Lundblad, C. (2007) "The risk return tradeoff in the long run: 1836–2003," *Journal* *of Financial Economics*, Vol. 85, No. 1, pp. 123–150.

Merton, R. C. (1973) "An intertemporal capital asset pricing model," *Econometrica: Journal of the Econometric Society*, pp. 867–887.

Mishra, S., L. Su, and A. Ullah (2010) "Semiparametric estimator of time series conditional Variance," *Journal of Business & Economic Statistics*, Vol. 28, No. 2, pp. 256–274.

Nelson, D. B. (1991) "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica: Journal of the Econometric Society*, pp. 347–370.

Pagan, A. R. and Y. Hong (1991) "Nonparametric estimation and the risk premium," *Nonparametric and Semiparametric Methods in Econometrics and Statistics, W. Barnett, J. Powellm and GE Tauchen (eds)*, pp. 51–75.

Pagan, A. R. and G. W. Schwert (1990) "Alternative models for conditional stock volatility," *Journal of Econometrics*, Vol. 45, No. 1, pp. 267–290.

Sun, Y. and Q. Li (2006) "An alternative series based consistent model specification test," *Economics Letters*, Vol. 93, No. 1, pp. 37–44.

Theodossiou, P. and U. Lee (1995) "Relationship between volatility and expected returns across international stock markets," *Journal of Business Finance & Accounting*, Vol. 22, No. 2, pp. 289–300.

Whitelaw, R. F. (2000) "Stock market risk and return: An equilibrium approach," *Review of Financial Studies*, Vol. 13, No. 3, pp. 521–547.

Wood, S. (2006) *Generalized additive models: an introduction with R*: CRC press.

Yang, L. (2006) "A semiparametric GARCH model for foreign exchange volatility," *Journal of Econometrics*, Vol. 130, No. 2, pp. 365–384.

Zhang, X., H. Wong, Y. Li, and W.-C. Ip (2013) "An Alternative GARCH-in-Mean Model: Structure and Estimation," *Communications in Statistics-Theory and Methods*, Vol. 42, No. 10, pp. 1821–1838.

APPENDIX A

FIGURES

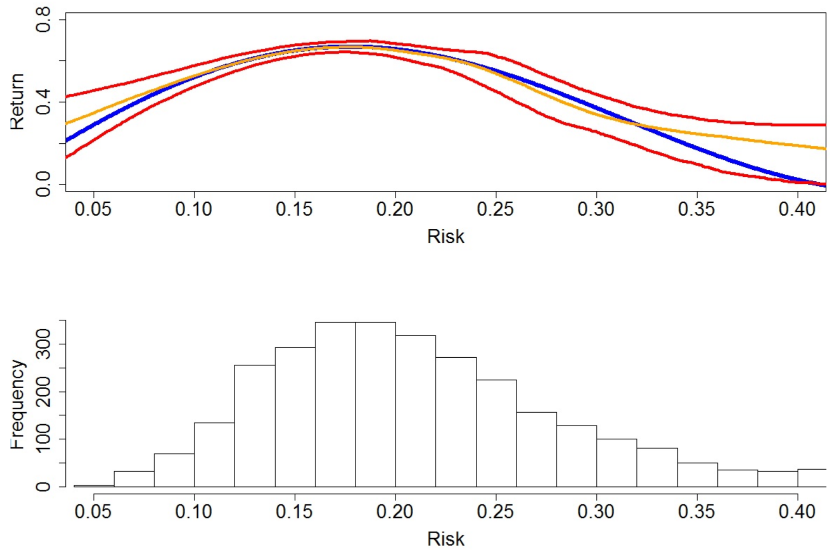

Figure A.1: Semi-parametric estimates of conditional mean equation for case N1

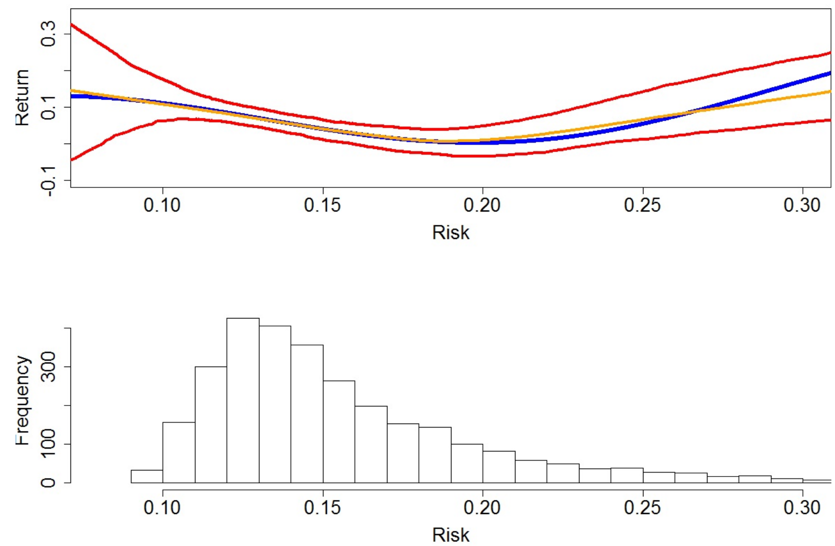

Figure A.2: Semi-parametric estimates of conditional mean equation for case N2

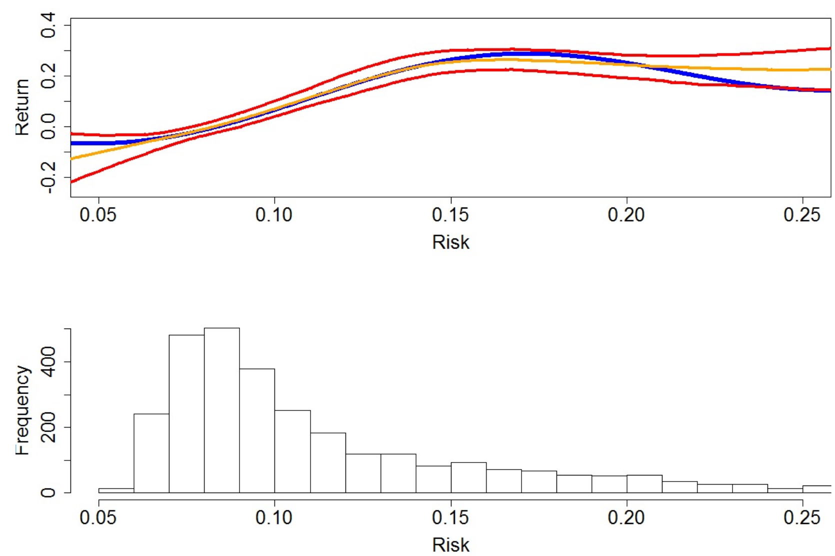

Figure A.3: Semi-parametric estimates of conditional mean equation for case N3

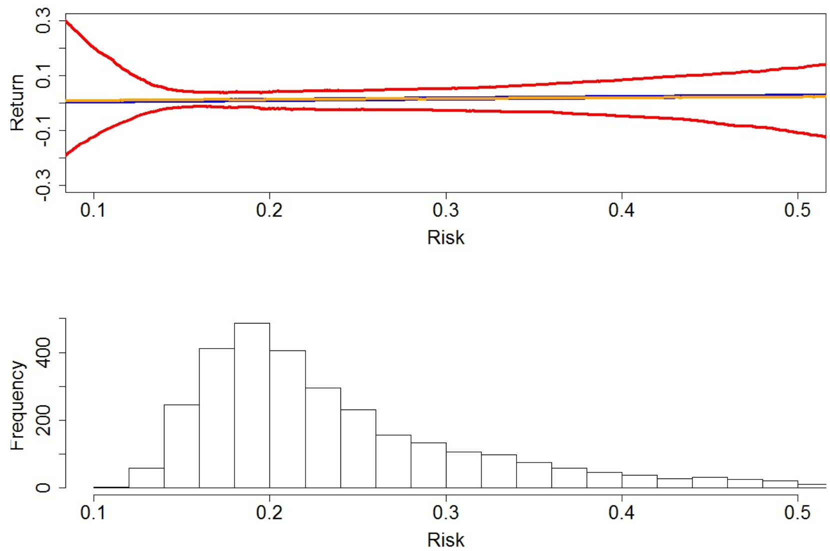

Figure A.4: Semi-parametric estimates of conditional mean equation for case L1

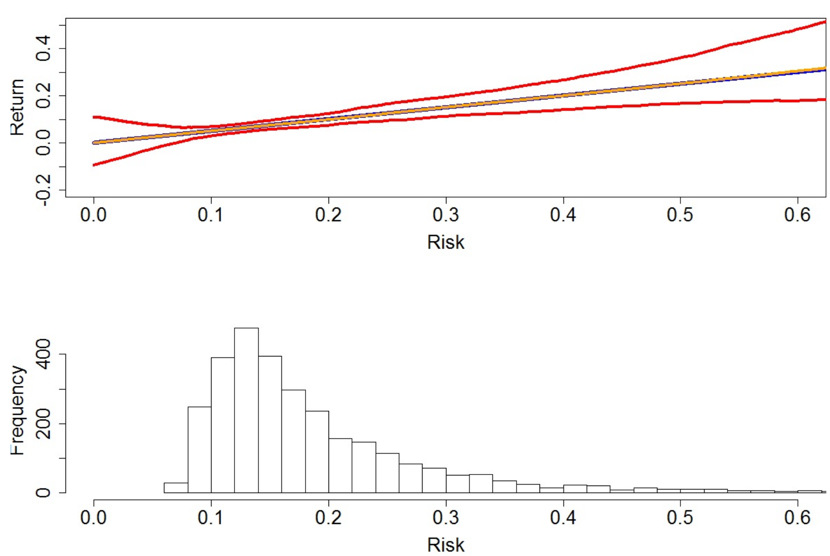

Figure A.5: Semi-parametric estimates of conditional mean equation for case L2

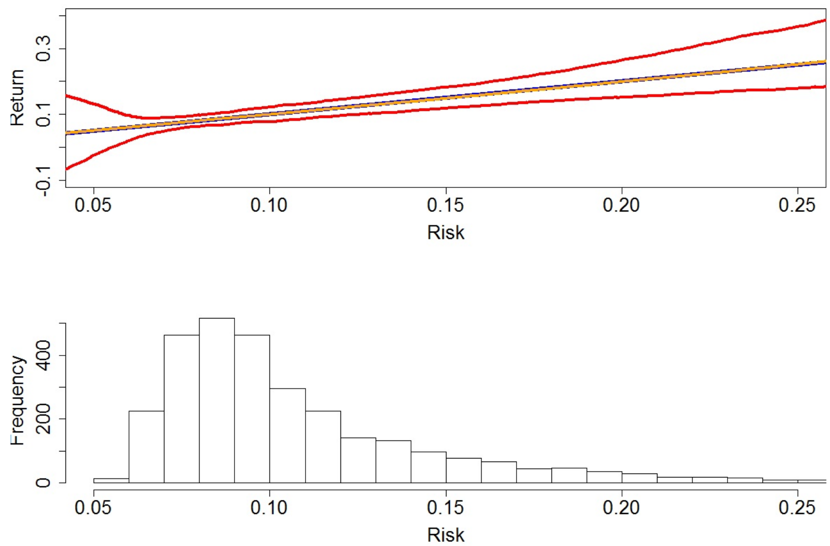

Figure A.6: Semi-parametric estimates of conditional mean equation for case L3

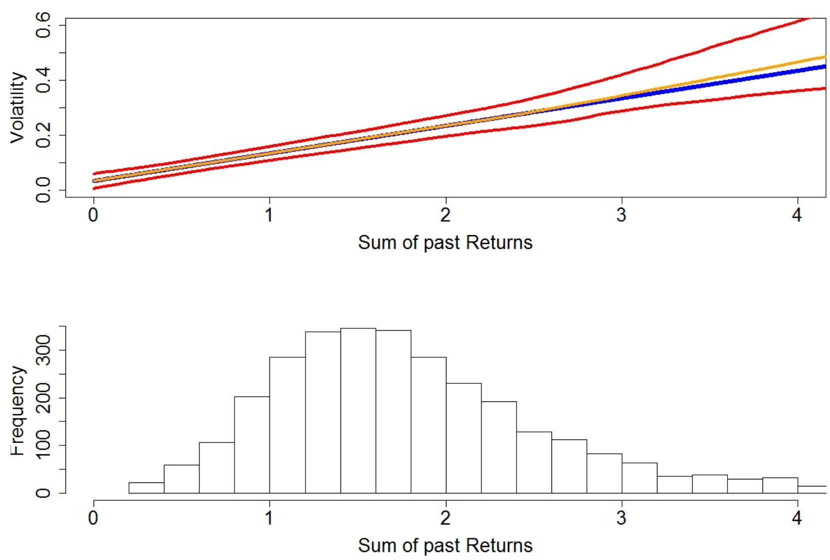

Figure A.7: Semi-parametric estimates of conditional variance equation for case N1

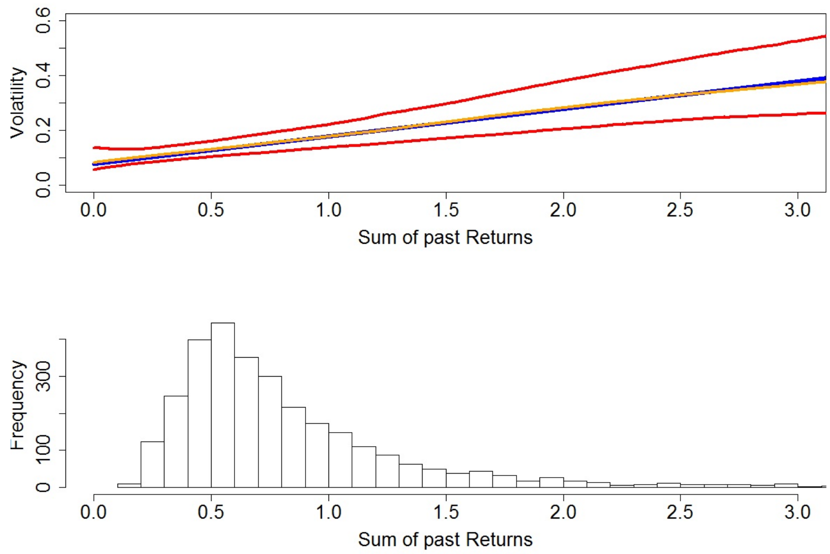

Figure A.8: Semi-parametric estimates of conditional variance equation for case N2

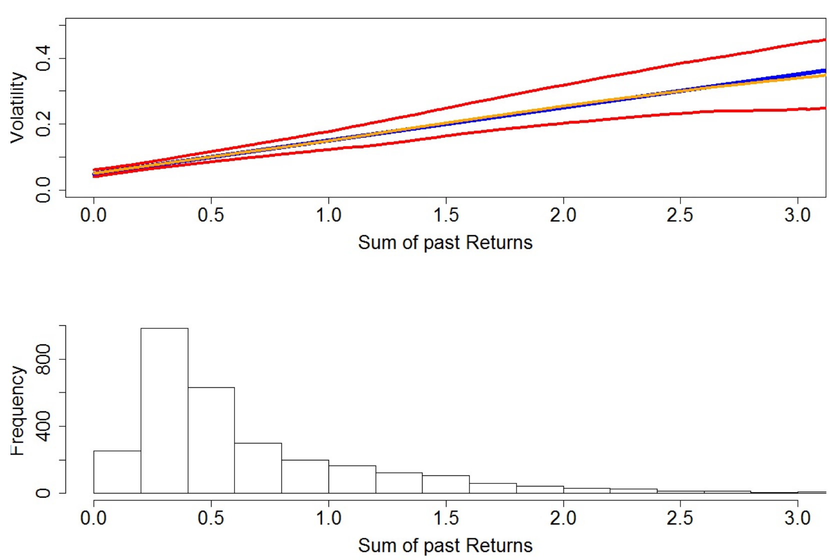

Figure A.9: Semi-parametric estimates of conditional variance equation for case N3

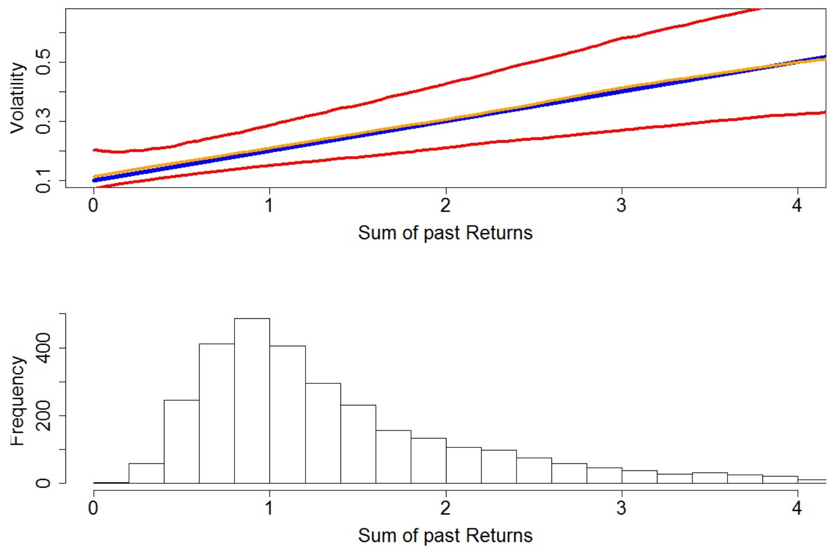

Figure A.10: Semi-parametric estimates of conditional variance equation for case L1

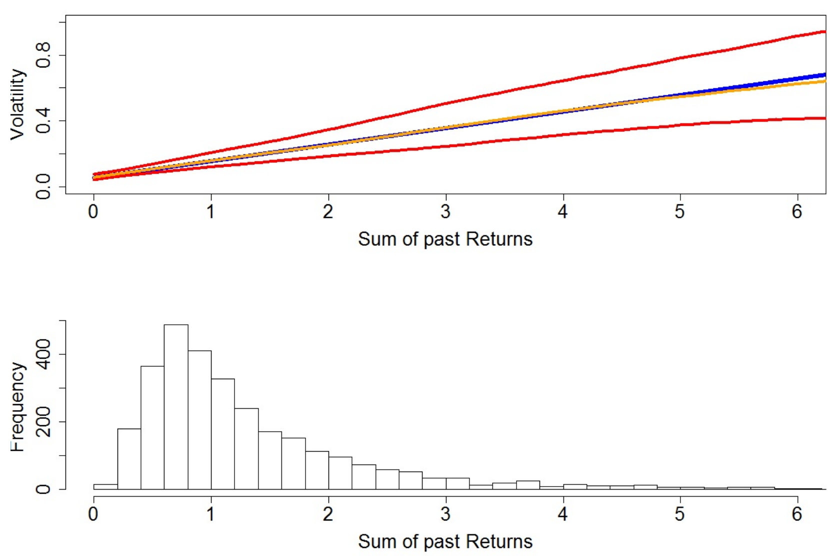

Figure A.11: Semi-parametric estimates of conditional variance equation for case L2

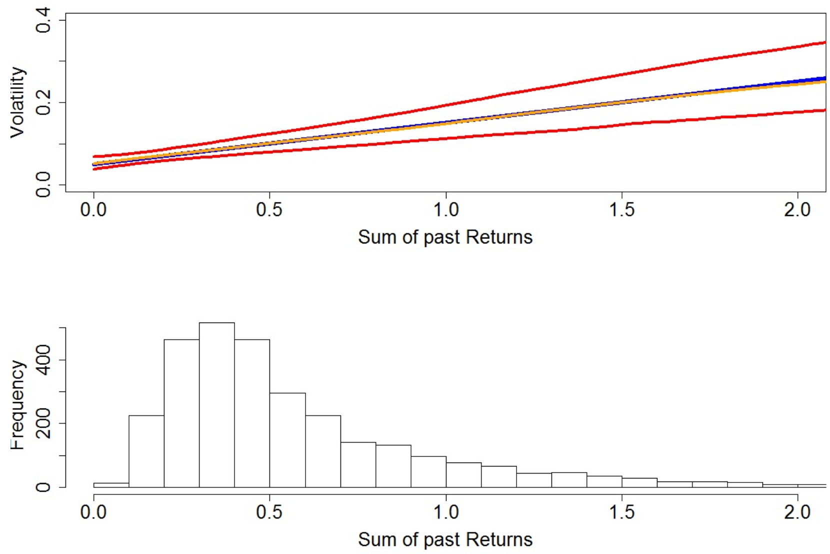

Figure A.12: Semi-parametric estimates of conditional variance equation for case L3

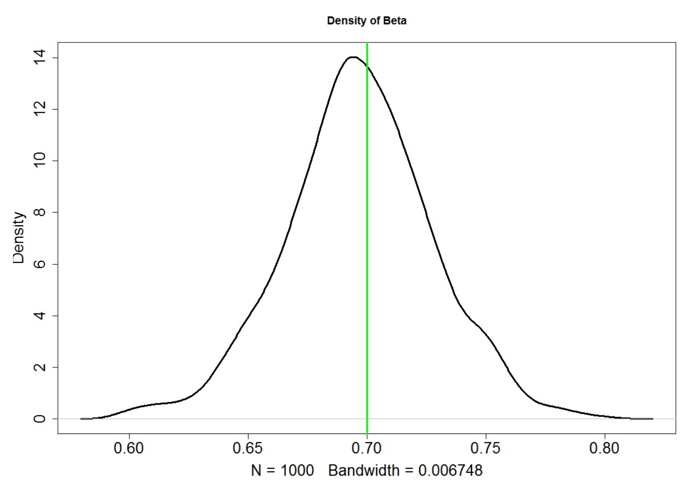

Figure A.13: Estimates of GARCH parameter, β, for case N1

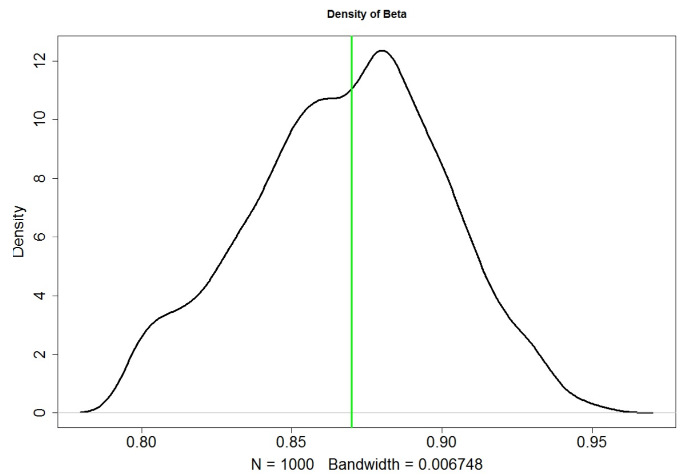

Figure A.14: Estimates of GARCH parameter, β, for case N2

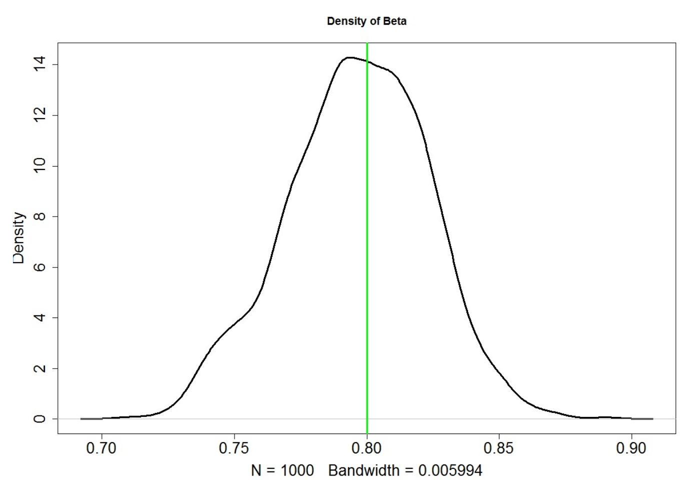

Figure A.15: Estimates of GARCH parameter, β, for case N3

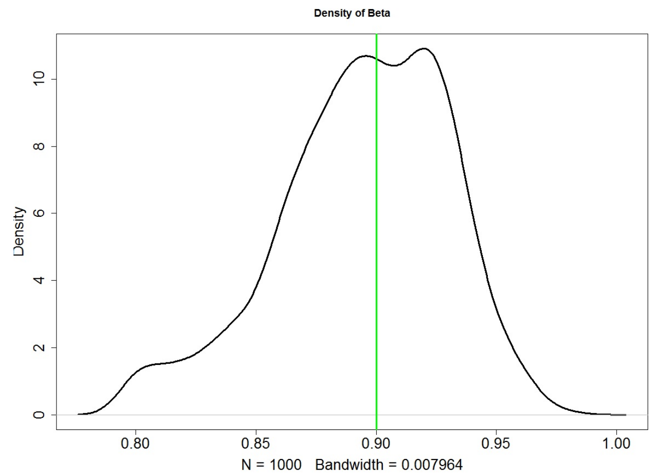

Figure A.16: Estimates of GARCH parameter, β, for case L1

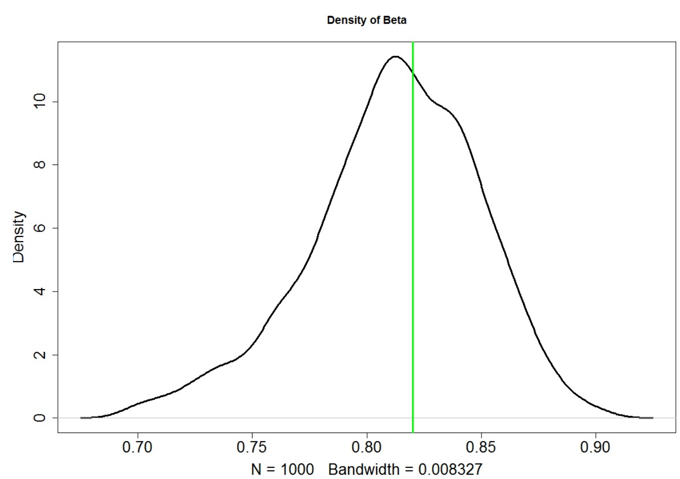

Figure A.17: Estimates of GARCH parameter, β, for case L2

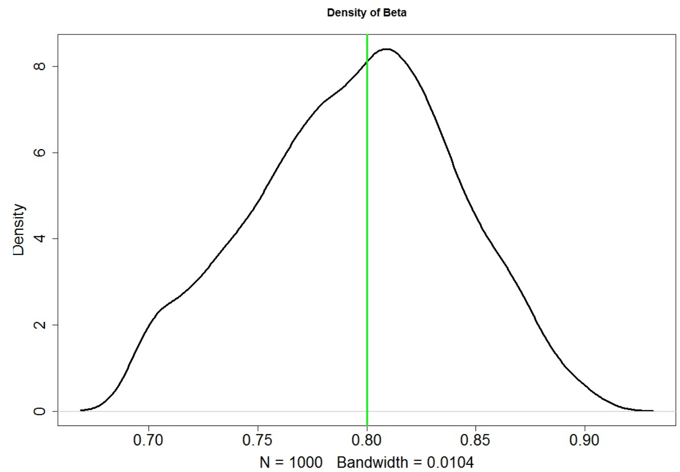

Figure A.18: Estimates of GARCH parameter, β, for case L3

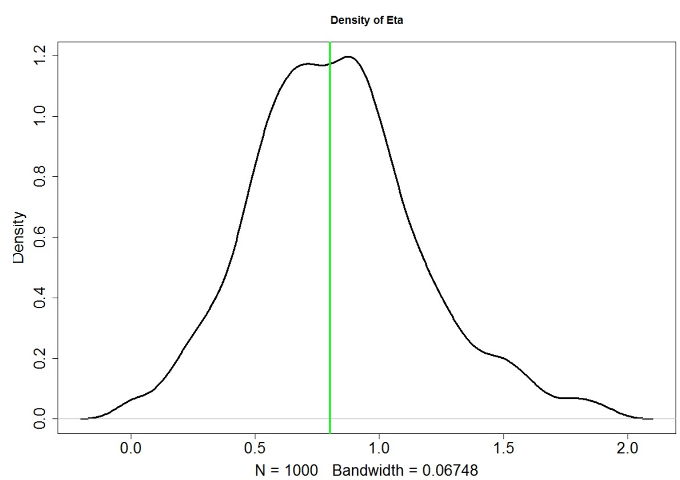

Figure A.19: N1: $\eta_0 = 0.80$

Figure A.20: Estimates of leverage effect parameter, η, for case N1

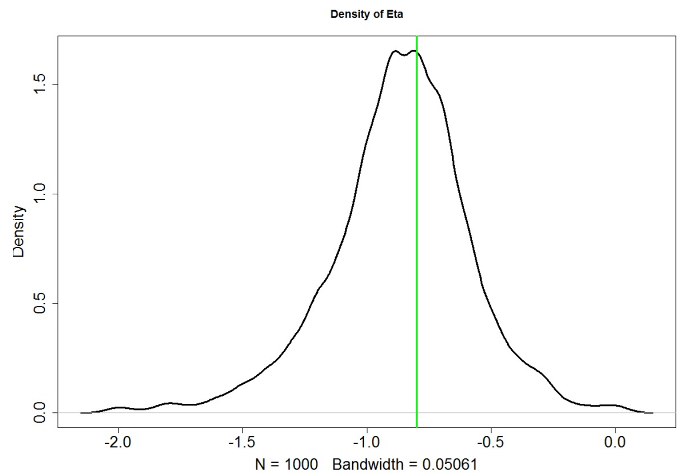

Figure A.21: Estimates of leverage effect parameter, η, for case N2

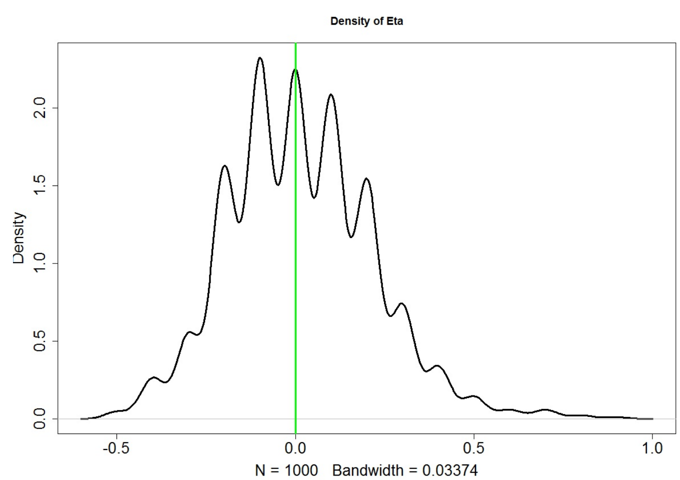

Figure A.22: Estimates of leverage effect parameter, η, for case N3

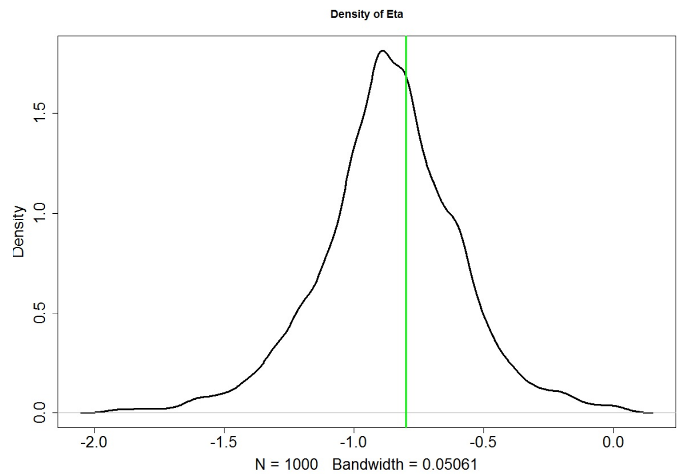

Figure A.23: Estimates of leverage effect parameter, η, for case L1

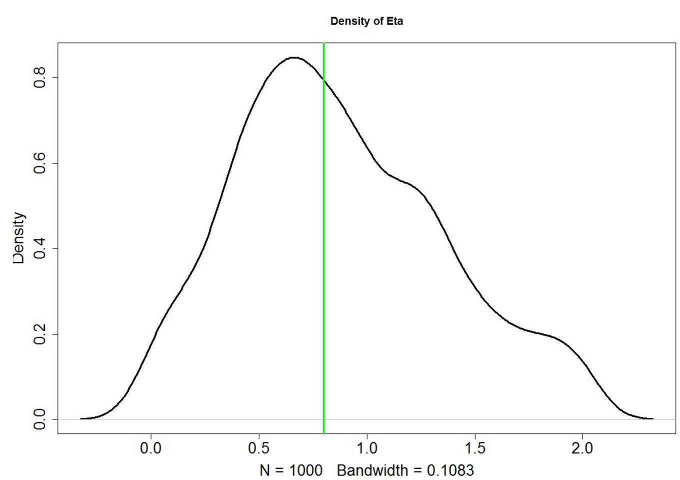

Figure A.24: Estimates of leverage effect parameter, η, for case L2

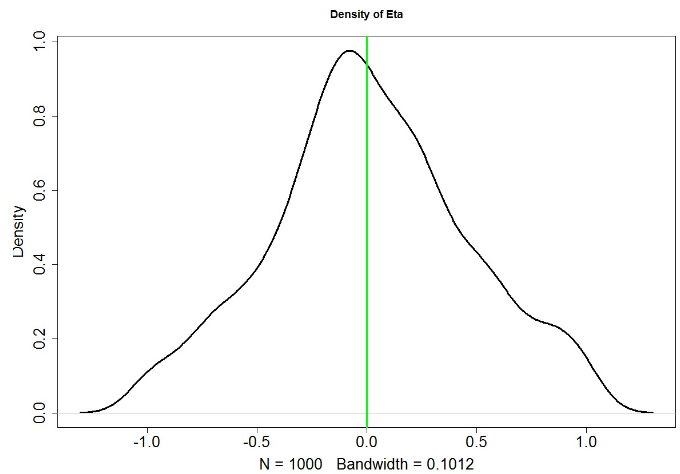

Figure A.25: Estimates of leverage effect parameter, η, for case L3

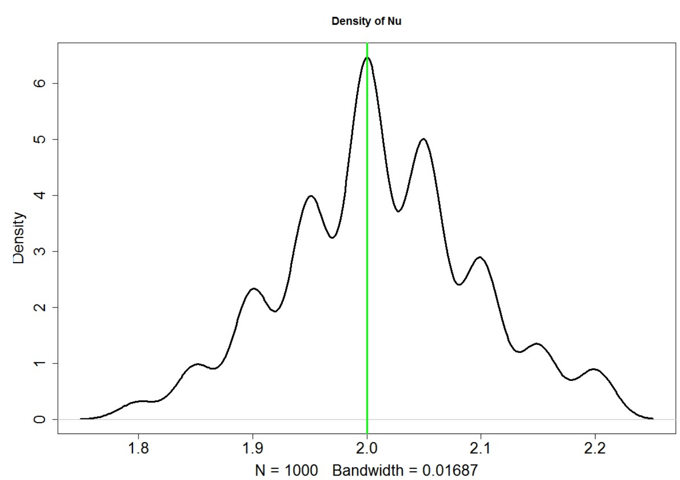

Figure A.26: Estimates of leptokurtic parameter, ν, for case N1

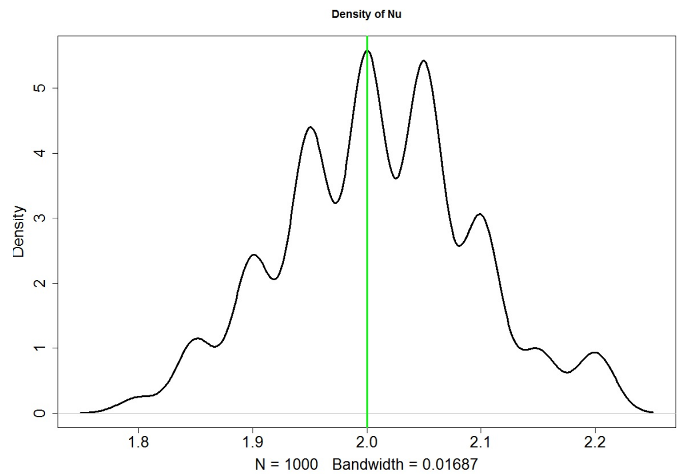

Figure A.27: Estimates of leptokurtic parameter, ν, for case N2

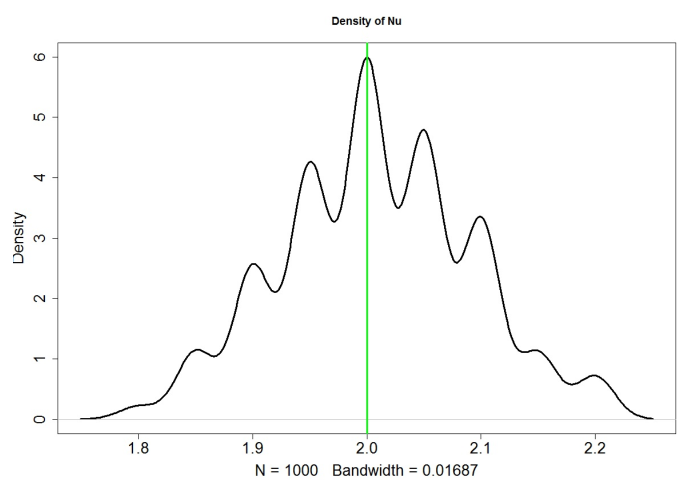

Figure A.28: Estimates of leptokurtic parameter, ν, for case N3

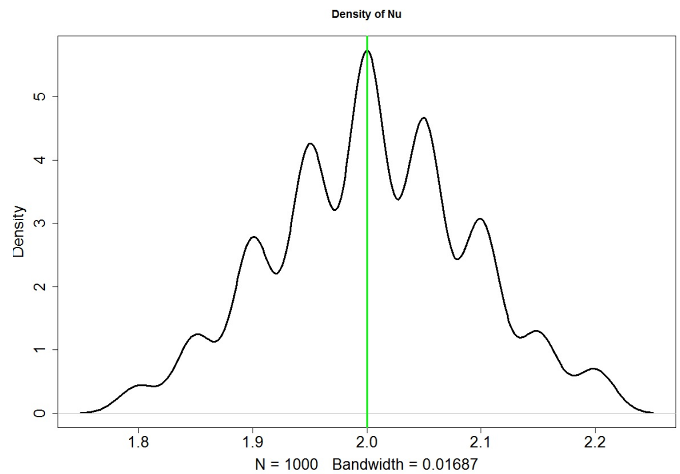

Figure A.29: Estimates of leptokurtic parameter, ν, for case L1

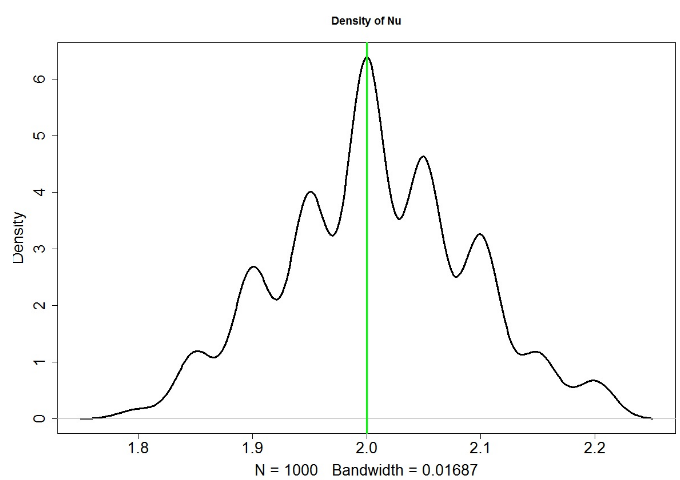

Figure A.30: Estimates of leptokurtic parameter, ν, for case L2

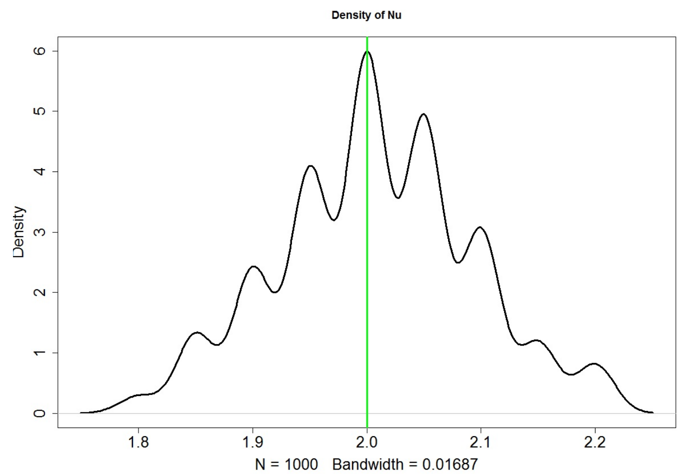

Figure A.31: Estimates of leptokurtic parameter, ν, for case L3

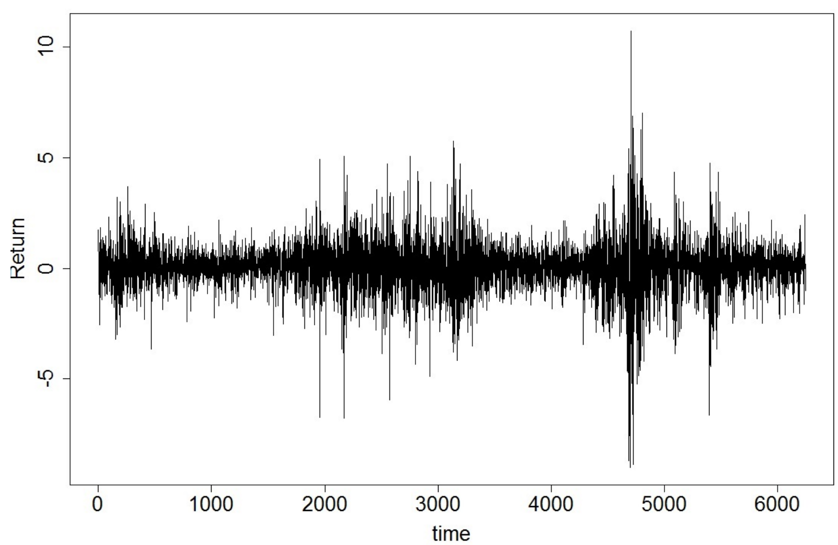

Figure A.32: Return of S&P 500 from 1990 to 2014.

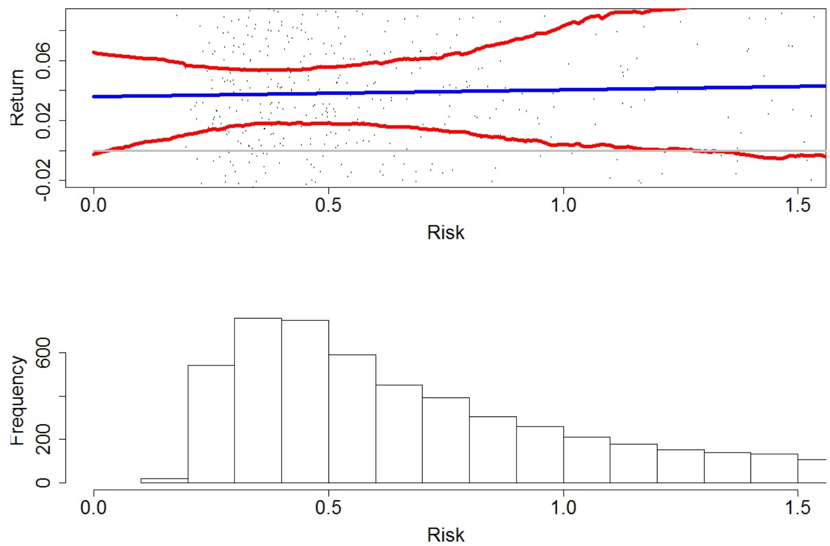

Figure A.33: Semi-parametric estimates of conditional mean equation

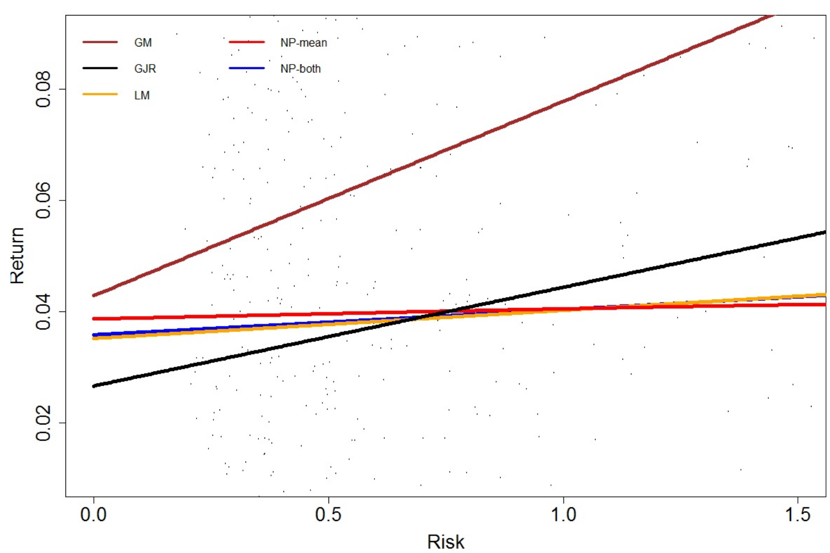

Figure A.34: Comparison of estimated mean equations

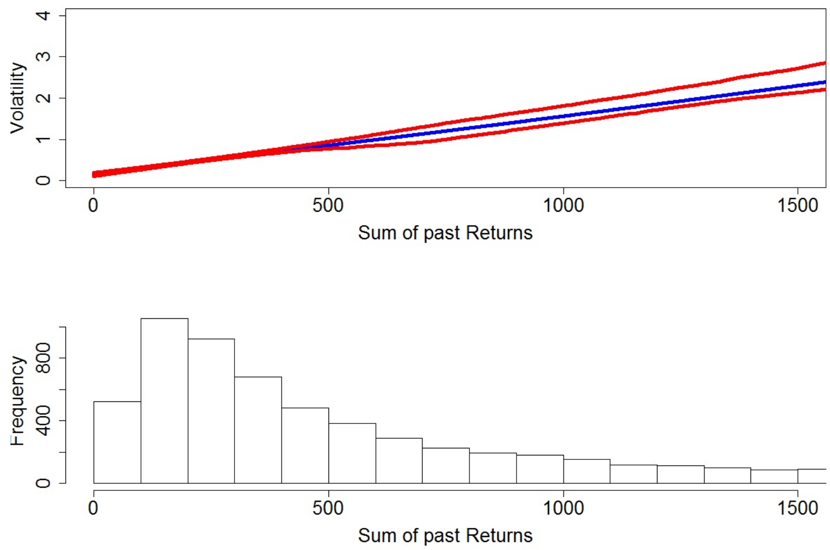

Figure A.35: Semi-parametric estimates of conditional variance equation

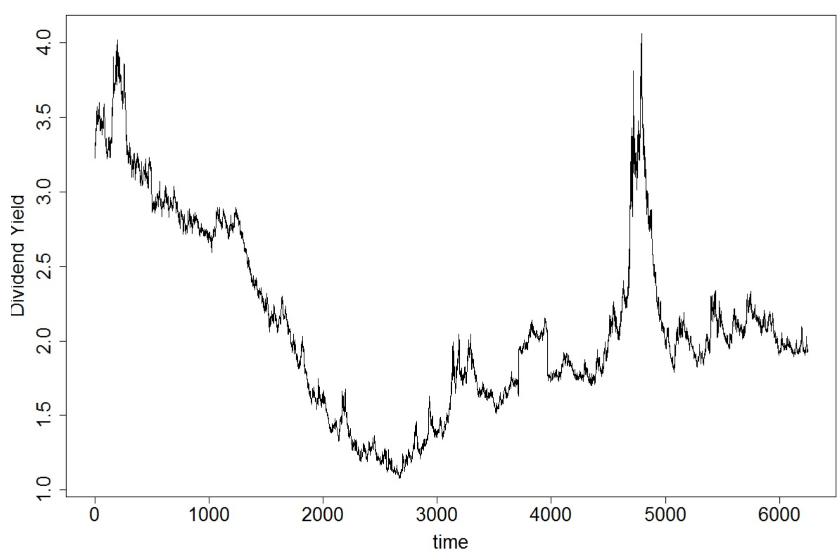

Figure A.36: Dividend yield

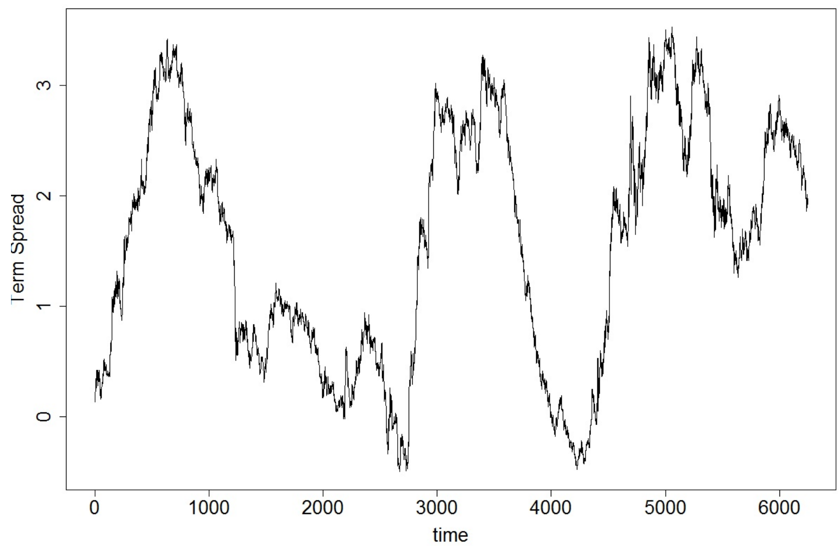

Figure A.37: Term spread

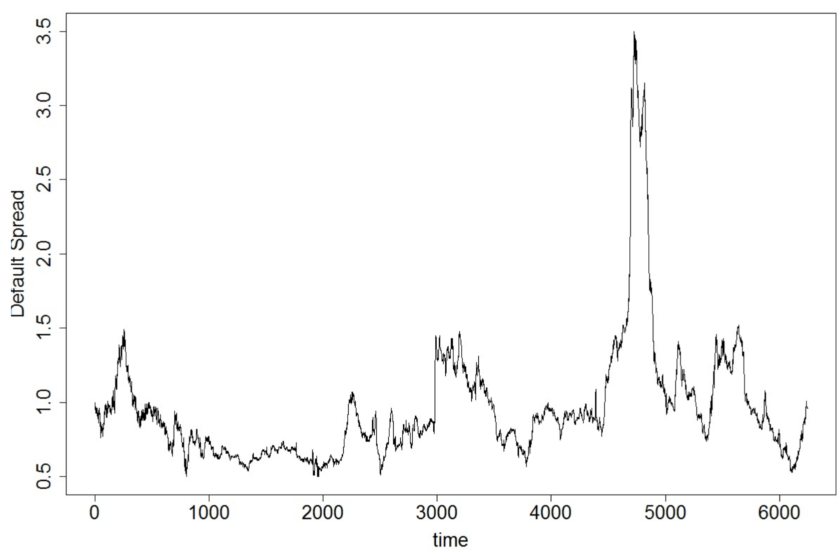

Figure A.38: Default spread

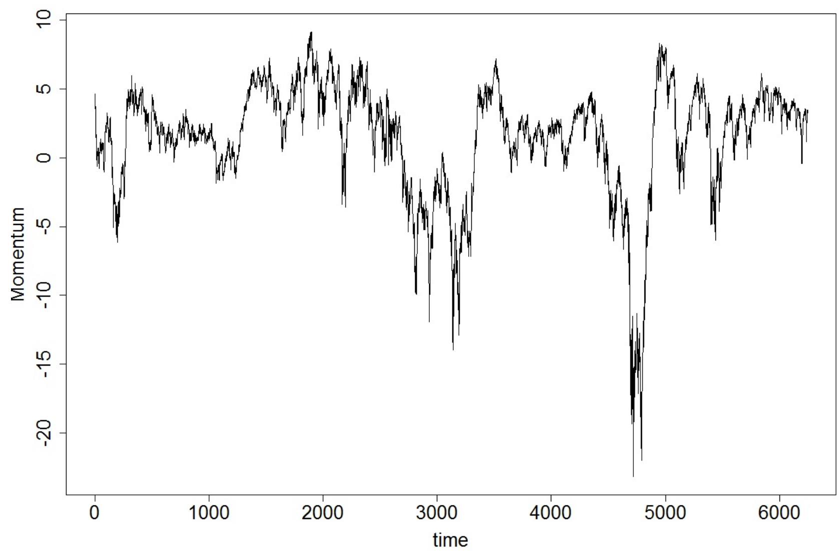

Figure A.39: Momentum

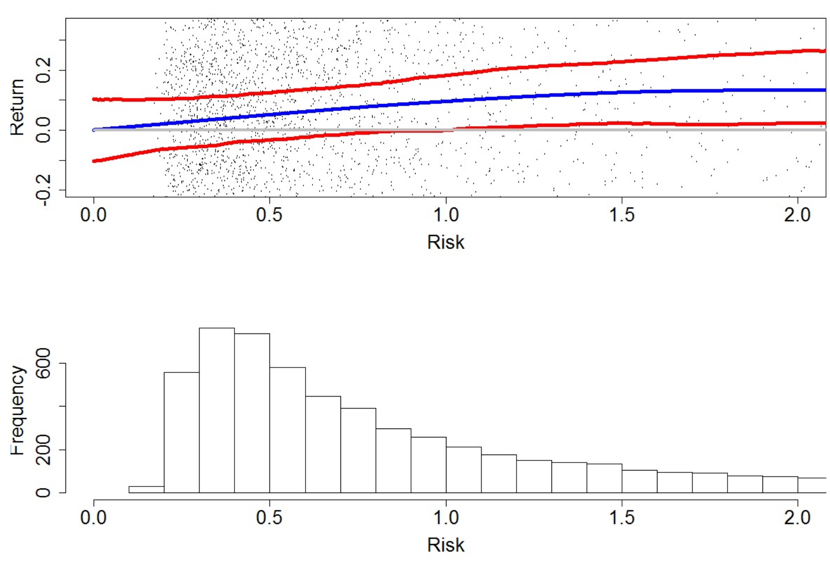

Figure A.40: Semi-parametric estimates of conditional mean equation

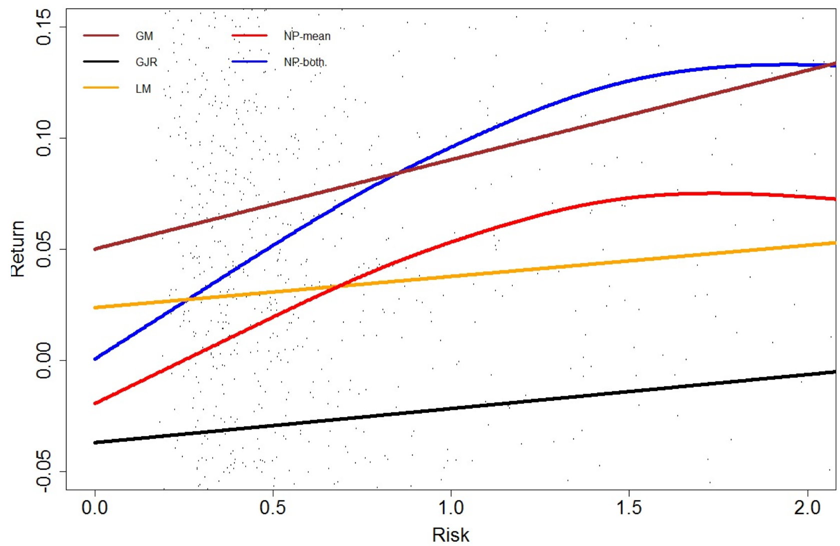

Figure A.41: Comparison of estimated mean equations

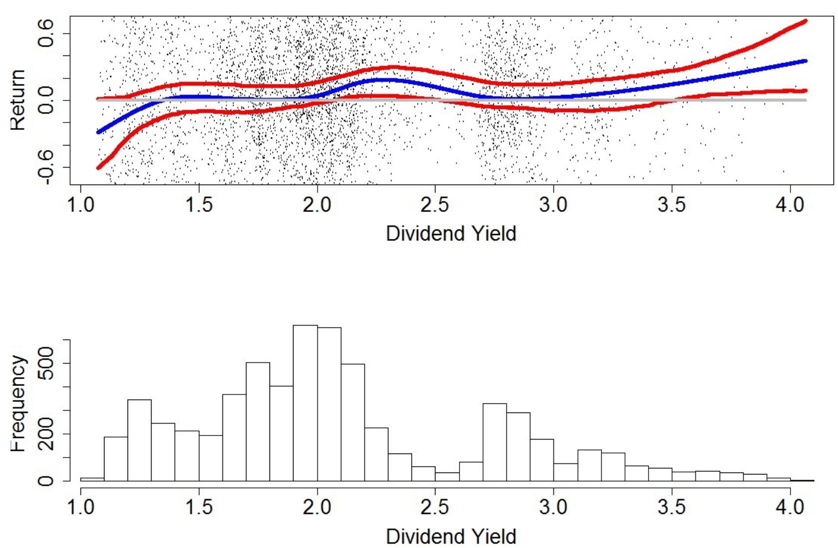

Figure A.42: The relationship between dividend yield and return

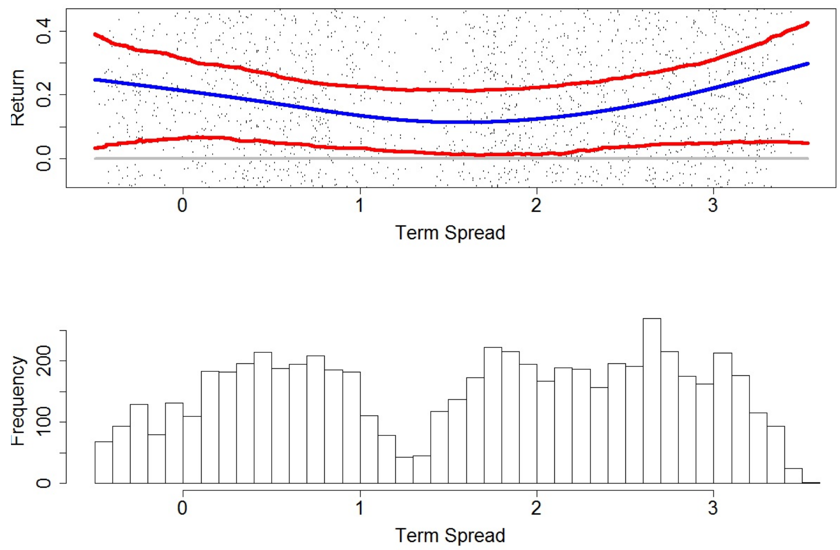

Figure A.43: The relationship between term spread and return

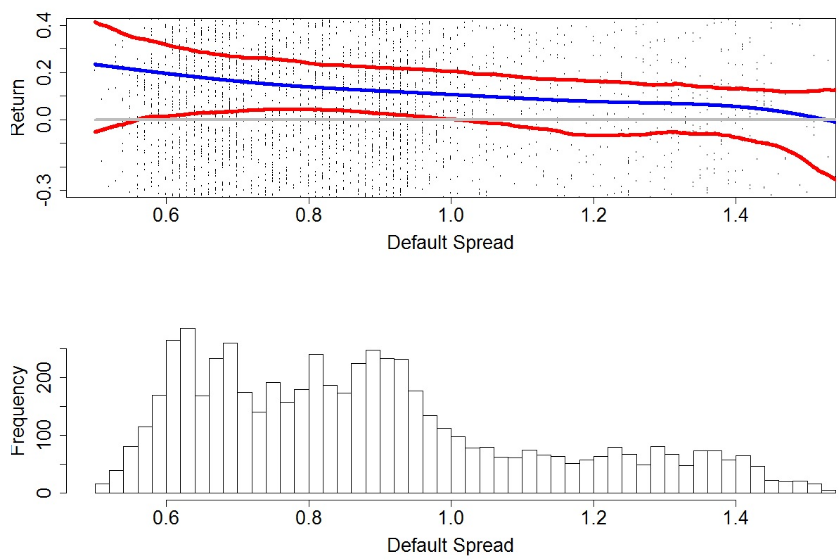

Figure A.44: The relationship between default spread and return

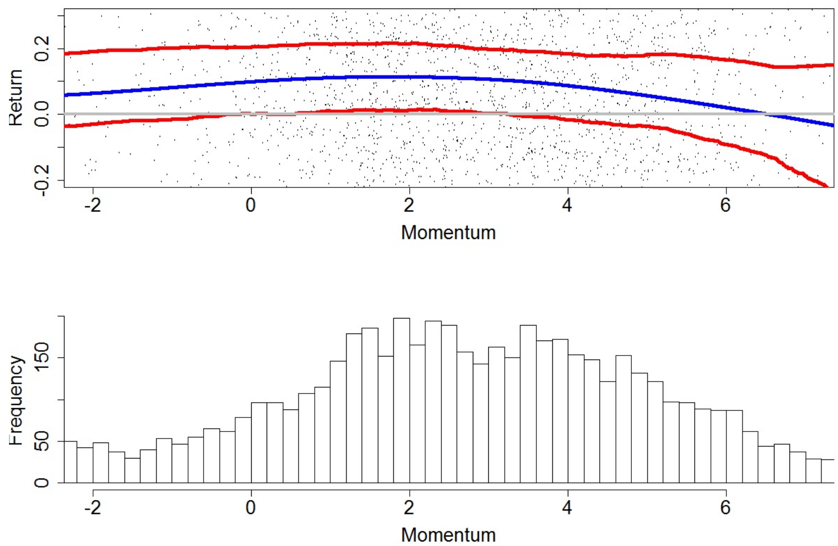

Figure A.45: The relationship between momentum and return

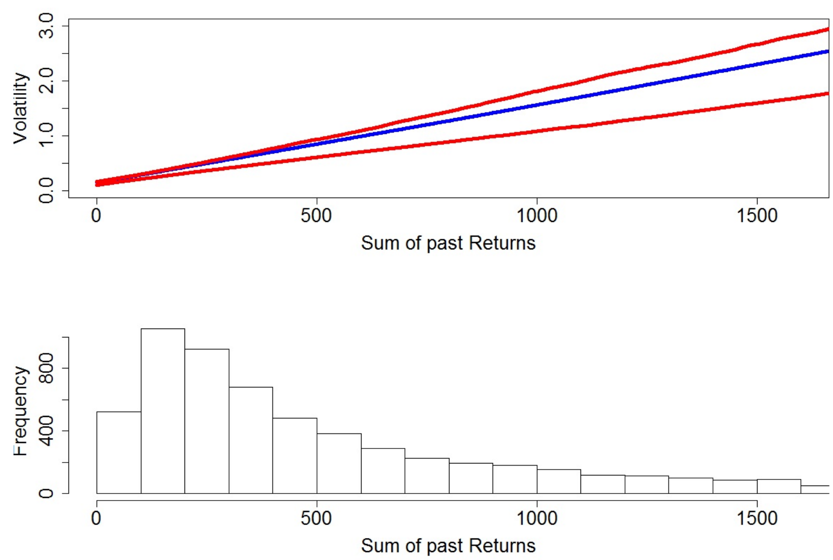

Figure A.46: Semi-parametric estimates of conditional variance equation

APPENDIX B

TABLES

Table B.1: Goodness of fit comparison for case N1

	MSE(mean)	MSE(var)
T=500	0.2209	0.003708
T=1000	0.2230	0.002456
T=2000	0.2242	0.001786
T=3000	0.2246	0.001636

Table B.2: Parameters estimation results I

	β_0	$Q5$	$\hat{\beta}$	$Q95$	η_0	$Q5$	$\hat{\eta}$	$Q95$	ν_0	$Q5$	$\hat{\nu}$	$Q95$
N1	0.70	0.65	0.70	0.75	0.8	0.3	0.8	1.5	2.0	1.85	2.0	2.15
N2	0.87	0.81	0.87	0.92	-0.8	-1.4	-0.9	-0.4	2.0	1.85	2.0	2.15
N3	0.80	0.75	0.80	0.84	0.0	-0.3	0.0	0.4	2.0	1.85	2.0	2.15
L1	0.90	0.83	0.90	0.94	-0.8	-1.3	-0.9	-0.4	2.0	1.85	2.0	2.15
L2	0.82	0.74	0.81	0.87	0.8	0.1	0.8	1.8	2.0	1.85	2.0	2.15
L3	0.80	0.71	0.80	0.87	0.0	-0.7	0.0	0.8	2.0	1.85	2.0	2.15

Table B.3: Parameters estimation results II

	β_0	mean	s.d.	η_0	mean	s.d.	ν_0	mean	s.d.
N1	0.70	0.6967	0.0311	0.8	0.8244	0.3434	2	2.0115	0.0849
N2	0.87	0.8683	0.0321	-0.8	-0.8680	0.2908	2	2.0095	0.0848
N3	0.80	0.7963	0.0267	0.0	0.0168	0.2050	2	2.0084	0.0838
L1	0.90	0.8944	0.0356	-0.8	-0.8624	0.2772	2	2.0045	0.0869
L2	0.82	0.8118	0.0377	0.8	0.8754	0.4846	2	2.0077	0.0832
L3	0.80	0.7948	0.0463	0.0	0.0156	0.4479	2	2.0080	0.0859

Table B.4: Comparison of estimated parameters for case N1

	β_0	$Q5$	$\hat{\beta}$	$Q95$	η_0	$Q5$	$\hat{\eta}$	$Q95$	ν_0	$Q5$	$\hat{\nu}$	$Q95$
T=500	0.70	0.60	0.69	0.79	0.8	0.0	0.8	2.0	2.0	1.80	2.1	2.20
T=1000	0.70	0.61	0.70	0.77	0.8	0.0	0.8	1.9	2.0	1.82	2.0	2.19
T=2000	0.70	0.63	0.70	0.76	0.8	0.2	0.8	1.6	2.0	1.83	2.0	2.18
T=3000	0.70	0.65	0.70	0.75	0.8	0.3	0.8	1.4	2.0	1.85	2.0	2.15

Table B.5: Specification test: rejection rate (%) at 5%

		L1	L2	L3	N1	N2	N3
T=500	Hong and White	2.1	2.1	2.1	75.8	5.8	49.1
	Sun and Li	8.8	3.9	4.1	82.8	14.1	51.6
T=1000	Hong and White	2.2	3.1	4.6	98.1	18.4	88.3
	Sun and Li	9.6	3.8	6.3	99.5	37.2	91.8
T=2000	Hong and White	2.7	3.5	5.9	100	51.9	99.2
	Sun and Li	5.9	2.7	5.3	100	67.3	99.8
T=3000	Hong and White	1.5	3.9	4.7	100	77.9	100
	Sun and Li	4.1	1.3	3.4	100	87.0	100

Table B.6: Parameters estimation results

Parameter	Estimation	Standard Error
β	0.92	0.0125
η	100	98.99
ν	1.4	0.0186

Table B.7: Goodness of fit: S&P 500

		1990-2014
MSE(mean)	GM	1.2842
	GJR	1.2788
	LM	1.2780
	NP-mean	1.2779
	NP-both	1.2780
MSE(var)	GM	12.4502
	GJR	12.1375
	LM	11.5769
	NP-mean	11.5750
	NP-both	11.5514

Table B.8: Parameters estimation results

Parameter	Estimation	Standard Error
β	0.92	0.0095
η	100	95.62
ν	1.4	0.0246

Table B.9: Goodness of fit: S&P 500 with 4 exogenous covariates

		1990-2014
MSE(mean)	GM	1.2841
	GJR	1.2804
	LM	1.2762
	NP-mean	1.2754
	NP-both	1.2613
MSE(var)	GM	12.4737
	GJR	12.1778
	LM	11.4996
	NP-mean	11.4520
	NP-both	10.8472