ABSTRACT

SAHA, PARAMITA. Robust Inference with Quantile Regression in Stochastic Volatility Models with application to Value at Risk calculation. (Under the direction of Professor Peter Bloomfield).

Stochastic Volatility (SV) models play an integral role in modeling time varying volatility, with widespread application in finance. Due to the absence of a closed form likelihood function, estimation is a challenging problem. In the presence of outliers, and the high kurtosis prevalent in financial data, robust estimation techniques are desirable. Also, in the context of risk assessment when the underlying model is SV, computing the one step ahead predictive return densities for Value at Risk (VaR) calculation entails a numerically indirect procedure. The Quantile Regression (QR) estimation is an increasingly important tool for analysis, which helps in fitting parsimonious models in lieu of full conditional distributions. We propose two methods (i) Regression Quantile Method of Moments (RQMM) and (ii) Regression Quantile - Kalman Filtering method (RQ-KF) based on the QR approach that can be used to obtain robust SV model parameter estimates as well as VaR estimates. The RQMM is a simulation based indirect inference procedure where auxiliary recursive quantile models are used, with gradients of the RQ objective function providing the moment conditions. This was motivated by the Efficient Method of Moments (EMM) approach used in SV model estimation and the Conditional Autoregressive Value at Risk (CAViaR) method. An optimal linear quantile model based on the underlying SV assumption is derived. This is used along with other CAViaR specifications for the auxiliary models. The RQ-KF is a computationally simplified procedure combining the QML and QR methodologies. Based on a recursive model under the SV framework, quantile estimates are produced by the Kalman filtering scheme and are further refined using the RQ objective function, yielding robust estimates.

For illustration purposes, comparison of the RQMM method with EMM under different data scenarios show that RQMM is stable under model misspecification, presence of outliers and heavy-tailedness. Comparison of the RQ-KF method with the existing QML method provide competitive results in terms of model estimation. Also, risk evaluation test results show desirable statistical properties of the quantile estimates obtained from these methods. Applications to real data and simulation studies on different parameter settings of the SV model provide empirical support in favor of the quantile model specifications.

We also propose an algorithm, based on a Gram Charlier density approximation for the conditional predictive volatility density given past returns, to compute the one step ahead predictive return densities in the existing Nonlinear Filtering (NF) scheme. This approach is used in likelihood and VaR computations. This algorithm provides an alternative approximation in the reduction of the infinite-dimensional state vector and is based on fewer computational steps compared to the existing methods. Results based on the algorithm are comparable to existing methods.

Robust Inference with Quantile Regression in Stochastic Volatility Models with application to Value at Risk calculation

by

Paramita Saha

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fullfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2008

APPROVED BY:

Dr. Denis Pelletier	Dr. Jason Osborne
Dr. Peter Bloomfield Chair of Advisory Committee	Dr. David Dickey

DEDICATION

To my parents and Suman.

BIOGRAPHY

Paramita Saha was born on October 17, 1978 in Calcutta, India. She graduated with a Bachelor of Science in Statistics from St. Xavier's College, University of Calcutta in September 2001 and a Master of Science degree in May, 2003 from Indian Institute of Technology, Kanpur. She joined the Department of Statistics at North Carolina State University as a graduate student in August, 2003. During her graduate study, she worked as a graduate industrial trainee at SAS Institute, Cary. After receiving her doctoral degree, she will join Novartis Oncology as a senior Biostatistician.

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to my advisor, Dr. Peter Bloomfield who has been a constant source of inspiration. His immense knowledge, insight and extensive support has always helped me and guided me. He has always motivated and encouraged me and I am extremely grateful to him for his patience. This dissertation would not have been possible without his guidance.

My sincere gratitude to Dr. Dickey, Dr. Osborne and Dr. Pelletier for giving me several suggestions during my oral Prelim exam and helping me throughout. I am grateful to them for being a part of the committee. It has been a pleasure and privilege being a part of their class.

I have been fortunate to be a part of the NCSU Statistics community. I am grateful to all my faculty members at NC State for their kind support. Their commitment and passion towards teaching is a source of great inspiration. I consider myself extremely fortunate at having received the guidance and encouragement of my dedicated teachers at NC State and IITK. I must also thank Adrian and Terry for being so very helpful and all my friends and family who have all been so supportive.

I also take this opportunity to thank Don Erdman, Kostas Kyriakoulis, Jan Chvosta, Terry Laurey and Jacques Rioux for providing such a wonderful and stimulating atmosphere at SAS. I must thank Don for being a great support throughout my GIT experience.

Without my parents' love, support and sacrifice, I would not have been able to reach where I am today. Under all circumstances, they have managed to give the best upbringing any child can ask for. I love you, Mom and Dad.

Life has never been the same since you came into my life, every moment spent has been special. We have had a great journey so far and we will rock on forever. My partner, my friend, my soulmate–Suman, I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In mathematical finance and financial statistics, the Stochastic Volatility (SV) model is used widely for modeling the time-varying volatility in financial markets. The development of the model and methodology related to its usage and growth is driven by a need for better pricing of options, efficient asset allocation and risk assessment. The scope of study in this thesis is related to risk assessment by calculating Value at Risk (VaR) and robust estimation using Quantile Regression (QR) techniques, when the underlying process is SV.

1.1 Motivation

A crucial aspect of measuring VaR is the volatility of the data. Hence, selecting a volatility model that fits the data appropriately, and its applicability, are important considerations in achieving the goal of VaR computation. In existing VaR methodology, much of the interest is concentrated towards the ARCH group of models introduced by Engle (1982), due to its deterministic nature of volatility and easier applicability. Relatively, the SV group of models are not as preferred, due to the volatility formulated as a stochastic process. With the emergence of high frequency data, SV models have again taken centerstage in econometric analysis, risk management and asset allocation. In pricing theory, continuous SV models have always formed the core of research. Hence, SV models form an attractive modeling tool with tremendous scope in multiple research arenas in finance.

An important feature of the SV models is that the volatility has its own stochastic process. This results in difficulties in the direct calculation of the likelihood and VaR estimation of SV models. In the 1990s, several computationally intensive methods were developed to calculate the likelihood numerically and efficiently.

Computation based methods such as Bayesian Monte Carlo (Jacquier *et al.*, 1994), Simulated Maximum Likelihood (SML) (Danielsson, 1994) and Simulated Method of Moments (SMM) (Duffie and Singleton, 1993), (Gourieroux *et al.*, 1993) methods produce better results in terms of accuracy and efficiency than the less computationally intensive tools such as Quasi Maximum Likelihood (QML) developed by Nelson (1988), Harvey *et al.* (1993), Ruiz (1994), Method of Moments (MM) technique (Taylor, 1986), Generalized Method of Moments (GMM) (Melino and Turnbull, 1990), (Andersen and Sørensen, 1996). Monte Carlo evidence suggests that MM, GMM, QML techniques suffer from poor small sample performance. Watanabe (1999)'s nonlinear filtering maximum likelihood (NFML) method provides a balance in that it gives good results in small samples without being as computationally demanding as a simulation based approach. Each of the aforementioned methods have their own set of advantages and we were motivated to study some of these methods (viz., QML, NFML, and EMM forming the basis for RQ-KF, NF scheme, RQMM methods, respectively) in the context of VaR calculation.

Calculation of VaR entails the knowledge of one-step ahead predictive densities. However, in the SV approach, the one step ahead densities cannot be obtained directly. Watanabe (1999) uses the QML estimates as a starting point for his nonlinear filtering (NF) method. The NF technique is an improvement over QML. Instead of using the first two moments as in the standard Kalman approach in QML, it uses the conditional densities directly in its filter to calculate the likelihood; hence, the conditional VaR's can be obtained subsequently. NF is based on a scheme that depends on a series of conditional integrals, where the integrals are evaluated numerically using piecewise linear approximations to the density function, motivated by Kitagawa (1987)'s work. One of the important considerations here is the choice of the number and location of the nodes that would suffice for a good approximation without compromising on computational demands. Appropriate nodes are chosen according to an approach suggested by Tanizaki (1996). The method computes the one-step ahead predictive densities as part of the algorithm. Hence, they can be used for likelihood computation (denoted by nonlinear filtering maximum likelihood (NFML)) and also as a filtering process for VaRs. NF's filtering method does not depend on the state space representation of the model, hence various extensions of the SV models can be easily tackled. Also, as long as known parametric forms of the volatility process and their relationship with their returns can be modeled, NF scheme finds ready application. Based on the aforementioned favorable attributes of the estimation method, it provides an

excellent groundwork for VaR calculation.

The Conditional Autoregressive Value at Risk (CAViaR) model (Engle and Manganelli, 2004) moves the focus of attention from the conditional distribution of returns directly to the behavior of the quantiles. The evolution of quantile over time is specified using an autoregressive quantile framework; hence the RQ objective criterion developed by Koenker and Bassett (1978) can be adapted for estimation. The Quantile Regression (QR) estimator is increasingly becoming an important tool for analysis, which helps in fitting parsimonious models in lieu of full conditional distributions. Another important aspect is that the method produces robust statistics and hence it finds its application in financial data scenarios with fat tailed distributions, and data with outliers. In applications, a robust inferential technique often proves to be beneficial in the case of misspecified models.

The motivation behind the methodologies proposed in this dissertation came from the CAViaR approach proposed by Engle and Manganelli (2004), based on the regression quantile framework introduced by Koenker and Bassett (1978). The CAViaR group of models provide a general framework for estimation of conditional VaRs. Some of their proposed models are derived from the popular autoregressive conditional heteroscedastic models. In ARCH type models, conditional quantiles are directly linked to the standard deviation of the distribution. Hence, an extension and application of their work in the SV context provides a strong motivation. This technique is particularly interesting because the likelihood computation in SV is circumvented which in itself involves an equally challenging problem.

The primary focus of research earlier was to efficiently and accurately estimate the SV parameters and volatility. However, consideration of heavy tailed distributions, and misspecification in models form an important premise in present day research. High kurtosis is endemic to financial data. We use this assumption heavily in our research. Asai (2008) compares an Asymmetric Stochastic Volatility (ASV) model with t distribution innovations with the multifactor SV (MFSV) model. Results using returns on the S&P 500 Composite and Tokyo stock price indexes and the Japan-US exchange rate indicate that the ASV-t model provides a better fit than the MFSV model on the basis of Akaike information criterion (AIC) and the Bayes information criterion (BIC). Several researchers, including Liesenfeld and Jung (2000), Watanabe and Asai (2003), show using a heavy tailed distribution for the error provide a better means in describing the high kurtosis in the data. The primary focus of this dissertation is to provide robust statistics for the estimation of

SV parameters and VaR calculation.

The efficient method of moments (EMM) technique (Bansal *et al.* (1993), Bansal *et al.* (1995), Gallant and Tauchen (1996)) is a flexible tool used for estimation. EMM is a culmination of the efficiency provided by the maximum likelihood (ML) approach coupled with the flexibility and affordability of the GMM. EMM is used where the ML method is infeasible or computationally intensive. Therefore, it finds its application directly in the SV context. The method employs an auxiliary data model that approximates the salient features of the true data generating process and has a readily computable likelihood function, in a closed form. The score equations of the auxiliary model are used as the moment equations to attain the efficiency of the ML asymptotically. The EMM as a method has its origins linked to the Indirect Inference (II), and Simulated Method of Moments (SMM) techniques.

The flexibility of the approach serves as the motivation behind our Regression Quantile Method of Moments (RQMM) methodology. It provides a general framework so that later developments can be incorporated easily. We use SV model as the benchmark throughout. The results obtained from our methodologies can be extended to the other situations like Asymmetric SV (ASV) models, Hermite Stochastic Volatility (HSV) models accounting for leverage effects, heavy-tailedness etc.

This dissertation addresses both robust statistical estimation and VaR calculation in a unified approach. Conditional quantiles or VaRs are related to the conditional distributions of the process. The Quantile Regression (QR) estimation is an important tool for analysis in fitting parsimonious quantile models in lieu of full conditional distributions. We propose two methods (i) Regression Quantile Method of Moments (RQMM) and (ii) Regression Quantile - Kalman Filtering method (RQ-KF) based on the QR approach that can be used for estimation of the SV model as well as for VaR calculation. The RQMM is a simulation-based indirect inference procedure where recursive quantile models are used as the auxiliary models, with the gradients of the RQ objective function (also known as the check or tick function) serving as the basis for estimating equations. This was motivated by the EMM approach used in SV model estimation and the CAViaR method proposed by Engle and Manganelli (2004). The check criterion gradient serves as a replacement of the score equations in EMM. The main concern is to develop robust statistics; therefore application of the check function as a nonparametric, flexible criterion provides a solution. Since GARCH is one of the preferred auxiliary models in EMM, the CAViaR models including

the iGARCH are reasonable choices. Hence, when the optimum parameter estimates for SV are attained, given that the auxiliary model approximates the true data model well, the estimated conditional quantiles will closely approximate the conditional quantiles of the SV, while also providing estimates for the SV model parameters. Details are provided in Chapter 3.

As an extension of CAViaR models, optimal linear quantile models (SVLIN and Asymmetric-SVLIN) based on the underlying SV assumption is derived. This is used along with the other CAViaR specifications as auxiliary models. The RQ-KF is a computationally simple procedure combining the good properties of QML and QR methodologies. The first step requires obtaining initial estimates using QML. Using a recursive model motivated by SV, quantile estimates are produced by the Kalman Filtering scheme based on the initial parameter estimates. These quantiles are then plugged into the check function to yield robust estimates. Details are provided in Chapter 4.

Financial data are known to exhibit heavy-tailedness characteristics. With simulation studies, these tools are shown to be stable in the presence of heavy-tailedness in the data. Comparison of the RQMM method with the EMM under different data scenarios show that RQMM is stable under local misspecification, and heavy-tailedness. Comparison of the RQ-KF method with the existing QML method provide competitive results in terms of model estimation. Also, risk evaluation test results show desirable statistical properties of the quantile estimates obtained from these methods. Applications to real data and simulation studies on different parameter settings of the SV model provide empirical support in favor of the quantile model specifications including the CAViaR, SVLIN and A-SVLIN.

Calculating the likelihood in SV model requires integration over an infinite dimensional state vector. Using Hermite polynomial approximations, we find an alternative finite dimensional approximation to the infinite dimensional state vector. We propose a Gram Charlier density approximation for the conditional predictive volatility density given past returns to compute the one step ahead predictive return densities in the existing Nonlinear Filtering (NF) scheme proposed by Watanabe (1999), Fridman and Harris (1998). Our method requires a reduced number of node points based on the coefficients of the Hermite polynomial, when compared to the earlier propositions. With this approximation, the conditional density calculated can be used in the likelihood function for SV parameters as part of the estimation process, or to find the conditional quantile for estimated parameters as part of the monitoring process. An algorithm based on a reduction to a finite-dimensional

state vector for evaluating with this approach is provided. The proposed algorithm can be used as a substitute for other numerical integration approximations based on choosing appropriate node points (see Watanabe (1999), Fridman and Harris (1998)).

Most of the earlier work for SV model estimation was coupled with obtaining a smoothed estimate of the volatility process by a nonlinear filter. In the Bayesian case of Jacquier *et al.* (1994) the smoothed estimates are obtained as a by-product of the method. In the non-Bayesian cases of MM and QML, standard approximate Kalman filtering schemes are used. Also, a quantile forecast in practice, is a two-step approach: model used to forecast volatility followed by a method of computing quantiles from volatility forecasts. Both our methodologies are aimed at parameter estimation and computing quantiles directly.

1.2 Outline of thesis

In Chapter 2, we give a brief overview of VaR and the existing methods for VaR calculation and evaluation. CAViaR model estimation by Regression quantiles is outlined briefly. We propose two linear quantile specifications based on the underlying assumption of SV. We compare the performance of these two models with the CAViaR models by using a real data analysis based on six stocks, and simulation studies based on several parameter settings of the SV model. The NF scheme is used as a benchmark to provide VaR estimates.

In Chapter 3, we provide a brief summary of EMM as our motivation. Next, we propose the RQMM methodology, followed by a discussion on the asymptotic properties of its estimates. We compare the performance of the RQMM with EMM under several heavy-tailed distributional assumptions, and misspecification with a simulation study. Under correct specification, the efficiency of RQMM estimates with EMM are also compared. It is to be noted that the VaR estimates using this method are reflective of our findings in Chapter 2. This is followed by a discussion.

In Chapter 4, we propose the RQ-KF method. We compare the performance of the method with QML with the help of a simulation study followed by a discussion.

In Chapter 5, we develop an algorithm based on the GC representation of the predicted conditional volatility density given the past returns. The incorporation of this scheme to the existing NF scheme is discussed. The application of such a scheme with respect to model estimation and VaR calculation is validated by a data exercise.

Chapter 6 summarizes the findings of the methods in Chapter 2–5 and discusses interesting future topics.

The remaining part of this chapter provides an overview of the volatility models.

1.3 Volatility models

1.3.1 SV models

The volatility of a financial asset is the standard deviation per unit time of the returns of an asset. In financial markets, volatility is a predominant feature, hence it plays a very important role in the determination of risk and valuation of options and derivatives. In financial terminology, volatility is the standard deviation and is directly related to the risk associated with holding financial securities, portfolio choice and investment decisions. The link between finance and Brownian motion goes way back till Bachelier (1900) who proposed a model for the French stock prices. A natural extension of a more reasonable model was proposed by Osborne (1959) where the price followed an exponential (or geometric) Brownian motion. The standard model that describes the behavior of the price process $P(t)$ is the solution of the stochastic differential equation (SDE) given by (1.1). Let us denote by $P(t)$ the continuous time process and let P_t denote its discrete analogue.

$$dP(t) = P(t)(\mu dt + \sigma dZ_1(t)) \tag{1.1}$$

where t is measured in units of one year, $Z_1(t)$ is a Brownian motion while the mean, μ and volatility, σ, are constant parameters of the model. The time convention is chosen to ensure that σ can be interpreted as an annualized volatility. This SDE has solution

$$P(t) = P(0)\exp\{\sigma Z_1(t) + (\mu - \frac{1}{2}\sigma^2)t\}. \tag{1.2}$$

The discrete time analogue of (1.1) based on a daily sequence of observations $(P_t)_{t\geq 0}$ is

$$lnP_t - lnP_{t-1} \equiv \Delta(lnP_t) = \nu + \sigma Z_t \tag{1.3}$$

where (Z_t) is a sequence of independent normal random variables with zero mean and variance 1/250 (number of business days per year).

Stochastic volatility models are useful because they try to incorporate the empirical observation that volatility appears to be stochastic. The SV candidate models have been motivated by intuition, convenience and tractability and some of them have been listed in equations (1.4), (1.5), (1.6), (1.7), and (1.8). Scott (1987), Wiggins (1987), Hull and White (1987), Hull and White (1988), Stein and Stein (1991) and Heston (1993) have each

proposed models of the form

$$\frac{dP(t)}{P(t)} = \sigma(t)dZ_1(t) + \mu dt \tag{1.4}$$

where $\sigma(t)$, the SV process is itself the solution of a stochastic differential equation. It is to be noted that in each of the cases enumerated hereunder, Z_2 is a Brownian motion perhaps correlated with the Brownian motion Z_1 which forms part of the specification of (1.4). Let us denote the correlation by ρ such that $E(dZ_1(t)dZ_2(t)) = \rho dt$. Further, we assume that ρ is a constant with modulus less than one.

$$d\sigma(t) = \sigma(t)(\alpha_0 dt + \gamma dZ_2(t)) \tag{1.5}$$

The first model given by (1.5) was introduced by Hull and White (1987) with $\rho = 0$ and Wiggins (1987) considered the general case. Here the logarithm of the volatility is a drifting Brownian motion. Scott (1987) considered the model (1.6) where the logarithm of the volatility is an Ornstein-Uhlenbeck (OU) process, and the discrete time analogue of the OU process is an AR(1) time series. The discrete version of this model is referred to as the SV model in the thesis, and we focus our work on this model. These models (1.5) and (1.6) have been formulated such that the volatility is positive.

$$d\ln\sigma^2(t) = \alpha_1(\alpha_0 - \ln\sigma^2(t))dt + \gamma dZ_2(t) \tag{1.6}$$

The model given below (1.7) was introduced by Scott (1987) and further investigated by Stein and Stein (1991) keeping $\rho = 0$. In this case, the volatility process itself is the OU process. However, the disadvantage of this model is that the volatility could easily become negative but (1.4) remains well defined.

$$d\sigma(t) = \alpha_1(\alpha_0 - \sigma(t))dt + \gamma dZ_2(t) \tag{1.7}$$

In 1988, Hull and White (1988) proposed a model and Heston (1993) investigated further with the general case of $\rho \neq 0$ of the form

$$d\sigma(t)^2 = (\alpha_0 - \alpha_1\sigma(t)^2)dt + \gamma\sigma_t dZ_2(t) \tag{1.8}$$

Two other models of note were proposed by Johnson and Shanno (1987) who modeled both the price and the volatility processes as constant elasticity of variance (CEV) processes (1.9) and another by Melino and Turnbull (1990) who assumed the price to follow a CEV process

(1.9) and the logarithm of volatility to be an OU process. The CEV model proposed by Cox and Ross (1976) is

$$\frac{dP(t)}{P(t)} = \sigma(P)dZ_1(t) + \mu dt \tag{1.9}$$

where $\sigma(P) = \sigma P^{\alpha-1}$ and $\alpha \in (0, 1)$. In CEV models, the price and volatility are negatively correlated.

Even though continuous time models provide the natural framework for analysis of option pricing, discrete time models are required for the statistical analysis for the calculation of VaR. The discrete time data models can be seen as a skeleton of the continuous time process. In the literature, the discretized version of the SV models have been termed as Stochastic variance models, autoregressive variance model etc., but in our discussion the SV model would mean the discretized SV model (1.6) (henceforth). In SV models, the logarithm of volatility is modeled as an AR(1) process with white noise. However, due to its non-deterministic nature, evaluation of the exact likelihood is challenging, hence they do not share the same popularity as the ARCH, GARCH models and have limited empirical applications.

In this dissertation, we use the SV model as defined in (1.10) below. Let y_t be the stochastic process of returns

$$y_t = \sigma_t \epsilon_t \quad \epsilon_t \sim N(0, 1) \tag{1.10}$$

where σ_t^2 is the conditional variance of the y_t. In the simplest SV model framework, the log of squared volatility is expressed as an AR model:

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + v_t \quad v_t \sim N(0, \sigma_v^2)$$

where ϵ_t and v_t are assumed to be independent of each other. The parameters of the model are denoted by $\theta = \{\alpha_0, \alpha_1, \sigma_v^2\}$.

Harvey and Shephard (1996) considered the leverage effect and introduced the Asymmetric Stochastic Volatility (ASV) model where an addition $\mathrm{Corr}(\epsilon_t, \mathrm{v_{t+1}}) = \rho$ was made to the SV model.

The Hermite SV (HSV) models were proposed by Meddahi (2001). This is a novel approach of modeling volatility in discrete and continuous time. The distinguishing feature in this model is that the variance is a linear combination of the Hermite polynomial functions of the state variable. Thus the dynamics of the volatility and the returns process

are governed by the dynamics of the state variables. This state variable may be governed by a Gaussian AR(1). It is important to note that the specification of the variance function based on the linear combination of the Hermite polynomials of state variable is a flexible model. Hence, the HSV model is formulated to keep the higher polynomial weights as free parameters that need to be estimated.

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \qquad \epsilon_t \sim N(0,1) \\ \sigma_t^2 &= a_0 + a_2(f_t^2 - 1) \\ f_t &= \beta f_{t-1} + \sqrt{1-\beta^2} v_t \qquad v_t \sim N(0,1) \end{aligned}$$

where $\theta = (a_0, a_2, \beta)$ are the parameters. The HSV models successfully generate fat tails for the variance and return processes.

For an overview of SV models, see Shephard (2005).

There are two main classes of discrete time models for stock prices with volatility. The first class, stochastic volatility models, is a discrete time approximation to the continuous time SV models that we outlined above. The second class constitutes the conditional heteroscedastic models.

1.3.2 Conditional Heteroscedastic models

The autoregressive conditional heteroscedastic (ARCH) model introduced by Engle (1982) is a pioneering work that led to the systematic development of a series of contributions that falls into the category of conditional heteroscedastic models. The econometric literature is replete with many models that are being used to quantify the uncertainty in future instantaneous volatility models. An ARCH (m) model is

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 {a_{t-1}}^2 + \ldots + \alpha_m {a_{t-m}}^2 \end{aligned}$$

where a_t is the mean corrected, serially uncorrelated asset return and ϵ_t is a sequence of i.i.d. random variables with mean 0 and variance 1. Another widely popular model that capture the dynamics of the volatility process is the GARCH structure given in (1.11). Bollerslev (1986)'s GARCH model is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \tag{1.11}$$

where $a_t = y_t - \mu_t$ is the mean corrected log return. Other extensions of the aforementioned models were made viz. EGARCH (Nelson, 1991) to capture asymmetry, and CHARMA (Tsay, 1987) which uses random coefficients to produce conditional heteroscedasticity. ARMACH(1,1) (Taylor, 1986) model is given by (1.12):

$$\sigma_t = \alpha_0 + \alpha_1 |a_{t-1}| + \beta_1 \sigma_{t-1} \tag{1.12}$$

EGARCH(m,s) (Nelson, 1991) model, shown in (1.13) is as follows:

$$\begin{aligned} \ln \sigma_t^2 = {} & \alpha_0(1 - \alpha_1 - \alpha_2 - \ldots - \alpha_m) + \\ & + \alpha_1 \ln \sigma_{t-1}^2 + \alpha_2 \ln \sigma_{t-2}^2 + \ldots + \alpha_m \ln \sigma_{t-m}^2 + \\ & + g(\epsilon_{t-1}) + \beta_1 g(\epsilon_{t-2}) + \ldots + \beta_{s-1} g(\epsilon_{t-s}) \end{aligned} \tag{1.13}$$

$$g(\epsilon_t) = \delta \epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)]$$

where δ and γ are real constants and the coefficients $(\delta+\gamma)$ and $(\delta-\gamma)$ show the asymmetry in response to the positive and negative values.

The GARCH model is a widely used tool to model financial data. Its strength lies in capturing volatility clustering of large price movements. The models described in this section are popular approaches to describe the changing volatility. The variance, σ_t^2, of the current return is written in terms of a nonstochastic function of the past observations. One of the many attractions of using such models is that the deterministic nature of the variance process leads to an exact likelihood, making estimation and forecasting straightforward.

Kim *et al.* (1998) provide evidence of better in-sample-fit of the SV model relative to GARCH-type models. Because of their well documented advantage in the literature, there is a need to develop and study methods that produce conditional quantile estimates, under the SV framework.

The next chapter gives a brief overview of the different methodologies used for Value at Risk calculation. The CAViaR model is discussed next followed by a proposition for SVLIN and ASVLIN models. These quantile specifications are derived from the relationship to the linear predictor of the latent volatility process under SV. Analysis of these models in the context of VaR performance are carried out with quantiles obtained from the NF scheme as benchmark, when the underlying data generating process is SV.

Chapter 2

VaR Estimation

With rising emphasis on risk management, portfolios are marked to market daily. The fluctuations intrinsic in the data captured, for instance, by volatility, play a very important role in VaR calculation. The two broad classes of modeling heteroscedasticity in financial data viz. stochastic volatility models and the ARCH group of models, provide a well established framework for VaR computation. Conditional heteroscedastic models such as ARCH, GARCH, EGARCH and others are often used to model volatility clustering, leverage effects etc. in the data. Stochastic volatility model is another widely used model with a wide application in finance. We focus this study on SV models. Due to the latent volatility modeled as a stochastic process, direct VaR computation is not straightforward.

The CAViaR group of models, proposed by Engle and Manganelli (2004), provides an unique way to directly estimate the conditional quantiles of interest, ie. Value at risk. This is achieved by minimizing the Regression Quantile (RQ) criterion, a robust loss function for a group of models based on conditional quantiles regression motivated by characteristics of financial data dynamics. Although the quantile specifications discussed under CAViaR, are motivated by the ARCH group especially, however, they can be applied more generally. Within this setup, we are interested in seeking a quantile model motivated by the SV model. We call this the SVLIN model. We further seek a quantile specification under the ASV model and name it the A-SVLIN model. With data generated from SV model, we evaluate the performance of the CAViaR models with the SVLIN model. The RQ criterion, devoid of any distributional assumptions, is used to gauge results in heavy-tailed error distribution situations, suitable for financial data. As a benchmark, we obtain the VaRs' directly from the nonlinear filtering, Watanabe (1999)'s method. We compare their performance based on

the Dynamic Quantile test (proposed by Engle and Manganelli (2004) and Chernozhukov (1999) independently) and other tests from the extant literature.

Section 2.2 gives a brief account of the existing methods followed by a description of the CAViaR method in the next section. Sections 2.4 and 2.5 discusses the RQ objective criterion introduced by Koenker and Bassett (1978), and tests for VAR evaluation, respectively. The SVLIN and A-SVLIN models are based on obtaining the best linear predictor using the Kalman filter for SV models and ASV models respectively, discussed in Section 2.6. The VaR computed from Watanabe (1999)'s method as benchmark is described in the Section 2.7. Results from the empirical study based on both application to stock data and a simulation study are presented in Section 2.8. The interpretation and results are discussed in Section 2.9.

2.1 Value at Risk

Value at Risk (VaR) is used as a standard measure for evaluating risk in financial institutions and organizations. It was introduced after the financial disasters of the 1990's. Although it started out as a tool to measure and monitor market risk, its usage has increased tremendously in the past few years to several other types of risk, eg, credit risk, liquidity risk, operational risk and subsequently to an integrated enterprise-wide risk. With globalization of financial markets leading to multiple sources of risk, pressure from regulators, and technological advances, its generalizability and implicit summarization of the risk scenario have led to its indispensable application to firm-wide risk management. Also, the boundaries between the abovementioned risks are becoming blurred. Hence, it provides an aggregate viewpoint for measuring a portfolio's risk. Its use is not only confined to derivatives but to all financial instruments. It provides an easy-to-use benchmark measure of risk, adopted by regulators (Basel committee on Banking supervision, U.S. Federal reserve, etc.), institutions with an exposure to financial risk, risk managers alike.

In recent years there has been an ever increasing demand to measure risk, and effective methods that can evaluate such critical situations are limited. The calculation of VaR has to be intrinsically related to the real world data scenarios. Dynamic models that closely replicate characteristics of financial data such as heavy tailedness, volatility clustering, and skewness have become relevant. In order to capture the time varying volatility, SV models are often used, in which the conditional variance is specified to follow a latent stochastic process. These models are mainly used in option pricing literature where stochastic

processes define the volatility equation. Their discrete counterparts are used to develop statistical methodologies for VaR calculation. Since VaR (or, quantiles) are directly related to the volatility process, they are tightly linked, and share similar characteristics. Due to the random component in the SV volatility equation, likelihood based model estimation, in itself poses a challenging problem, and so does VaR computation. They nevertheless provide a very attractive alternative to time varying conditional variance modeling, and it is of interest to study methods that compute VaR under this modeling framework.

Engle and Manganelli (2004) proposed autoregressive models for the VaRs, estimating the parameters by minimizing the Regression Quantile criterion (Koenker and Bassett, 1978). Four models were proposed which can be seen as extensions to the widely used conditional heteroscedasticity (ARCH) models (Bollerslev, 1986).

A natural extension to their work is to implement a similar objective of finding conditional VaR calculation methods in a regression setup, under the SV model framework. This proposition is especially attractive in the context of SV since this technique circumvents the likelihood calculation and directly yields VaR estimates. Also, their performance in the context of heavy tailed distributions should also be considered, since financial data exhibits high kurtosis. The empirical study in Section 2.8 verifies that the proposed model successfully describes the evolution of quantiles at the tails, especially, in cases of SV models with heavy tailed error distributions.

Let us denote $\Delta V_{t,l} = P_{t+l} - P_t$ as the change in the value of the financial position from time t to $t+l$, where P_t is the price process and $F_l(x)$ is the cumulative distribution function (cdf) of $\Delta V_{t,l}$. $-\mathrm{VaR}_t(\tau, l)$ can be defined as the loss faced by a financial position during a given time period $(t, t+l)$ for a given confidence level τ under normal market conditions. We can formally define VaR of a long position over a time horizon l with probability τ as

$$\tau = \Pr[\Delta V_{t,l} \leq -\mathrm{VaR}_t(\tau, l)] = F_l(-\mathrm{VaR_t}(\tau, \mathrm{l}))$$

where the VaR, $-\mathrm{VaR}_t(\tau, l)$ is a negative value (loss), by the above notation. Following convention, $\mathrm{VaR_t}$ is a positive value measured in the currency of interest and τ is the confidence level. Hence we would look at the entire quantity $-\mathrm{VaR_t}$ as Value at Risk (VaR) henceforth. Usually for analysis purposes, instead of working with values of assets we use

the log returns. Let us denote the log returns by

$$y_t = \ln \frac{P_t}{P_{t-1}}$$

Hence a negative y_t means loss. Modeling and estimation is usually carried out by using log returns because they have been shown to exhibit the properties of volatility clustering, stationarity, possessing almost zero autocorrelation, conditional heteroscedasticity. Hence, a VaR definition needs to be formalized with respect to the returns. Following the notation of Engle and Manganelli (2004), let $\{y_t\}_{t=1}^{T}$ be the time series of portfolio returns. VaR at a time t is given by

$$Pr[y_t < -\mathrm{VaR_t}|\Omega_{t-1}] = \tau$$

where Ω_{t-1} denotes the information set at the end of time $t-1$. Analyzing the left or the right tail of the c.d.f. depends on whether you are a holder of the long position or a short position respectively. The above definition is applied by the holder of a long position because he faces a loss when the value of the portfolio decreases. However, a change in the sign of the variable would make the holder of a short position study the left tail as well. Hence, it suffices to use this definition for the following discussion.

Some of the prevalent statistical methodologies frequently used in evaluating VaR can be broadly categorized into the following:-
Econometric approaches (EA), *quantile regression* (QR), *extreme value distributions* (EVD) and *historical simulation* (HS). $RiskMetrics^{TM}$, proposed by J.P. Morgan, is a particular case of EA and is a widely used tool. For an overview of VaR, see Jorion (2007). The next section sketches a brief review of these methods followed by an outline of the rest of the paper.

2.2 Existing Methods of VaR Estimation

J. P. Morgan developed $RiskMetrics^{TM}$ methodology for VaR calculation. For a better exposition, the reader is referred to Logerstaey and More (1995). This method assumes that the continuously compounded daily return of a portfolio follows a conditional normal distribution. RiskMetrics assume that $y_t|\Omega_{t-1} \sim N(\mu_t, \sigma_t^2)$, where μ_t is the conditional mean and σ_t^2 is the conditional variance of y_t. In addition, it also assumes that the two quantities evolve over time following the simple model:

$$\mu_t = 0, \qquad \sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)y_{t-1}^2, \qquad 0 < \alpha < 1.$$

Hence, it assumes that the daily log returns satisfying $y_t = \sigma_t \varepsilon_t$ is an IGARCH(1,1) process without a drift. For such a random-walk IGARCH model, the conditional distribution of a multiperiod return is easily available. Specifically, for a k-period horizon the conditional distribution $y_t[k]|\Omega_t$ is $N(0, \sigma_t^2[k])$, where

$$y_t[k] = \sum_{i=t+1}^{t+k} y_i$$

and $\sigma_t^2[k]$ can be shown to be equal to $k\sigma_{t+1}^2$. Therefore, for a k-period horizon, it follows that $VaR(k) = \sqrt{(k)} \times VaR$.

There are several econometric models for the mean and the volatility processes. To show an example, for a general time series model for y_t let us use the ARMA, GARCH to model the mean and the volatility processes respectively.

$$y_t = \phi_0 + \sum_{i=1}^{a} \phi_i y_{t-i} + a_t - \sum_{j=1}^{b} \theta_j a_{t-j}$$

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2$$

where a_t is the mean corrected, serially uncorrelated asset return and ϵ_t is a sequence of i.i.d. random variates with mean 0 and variance 1. The residuals may be assumed to follow a known parametric distribution (e.g., Normal, t etc.). Other conditional heteroscedastic models viz. IGARCH, GARCH-M, EGARCH, CHARMA, etc. can also be used to model the volatility process. Hence, quantiles from these conditional distributions can be evaluated accordingly. This method is relatively simple to use.

Quantile regression is a nonparametric approach to VaR computation. It is more appropriate to include the covariate information while estimating the quantiles for the conditional distributions. Hence, in this case, Ω_{t-1} introduced earlier which represent the information till the end of period $t-1$ contains further information on the covariates. Koenker and Bassett (1978) developed the quantile regression theory by formulating the sample quantile problem to a linear regression one. One of the recent papers using the regression quantile method is the CAViaR model proposed by Engle and Manganelli (2004). They propose an autoregression of the conditional VaRs' and their approach is discussed in detail, in Sections 2.3 and 2.4.

The generalized extreme value distribution (EVD) of Jenkinson (1955) encompasses the three types of limiting distribution of Gnedenko (1943). Since we are interested in the extreme tail of the distribution for the computation of VaR and because the EVD is appropriate for analyzing extreme events, EVD theory is often used in VaR calculations. The idea is to partition the sample of the returns into nonoverlapping subintervals. Using the minimum values from these subintervals, we estimate the EVD parameters. Estimation of the three parameters in the EVD can be carried out parametrically by maximum likelihood or by minimizing the sum of squared errors in a regression setup, or nonparametrically. Once the distribution is known, VaRs can be calculated. This method depends on the choice of the subintervals for a given data set. Hence, in order to make the minimum among the subinterval points appreciably close to the real extreme data, sufficient data are required.

A common method for VaR estimation is historical simulation (HS), in which the simulated distribution of returns is simply the empirical distribution of the past observations. The advantage of this method is that it makes no distributional assumptions, that it is nonparametric. If historical data has been collected in-house, the same data can be used for VaR calculation. It is relatively easy to implement; however, it has major drawbacks. The method applies equal weight to its past observations. Since it considers only one realization of the data, in cases of large deviations of the data from the true distribution, the quantiles obtained by HS can be greatly affected. Moreover, it is slow to capture structural breaks that can be easily detected by, for instance, Riskmetrics. The choice of sample size can also have a huge effect on the predicted values. Some variations have been proposed in the literature to overcome some of these disadvantages. For example, Boudoukh *et al.* (1998) proposed a method which applies exponentially declining weights to the past returns.

Although its conceptual simplicity has attracted a wide range of users, the computation of VaR is a challenging statistical problem and most of the methods developed so far are based on simplifying assumptions. Standard model free methods such as historical simulations, rely on a single sample path and does not require any distributional assumptions. Reliable risk measurement requires to account for the characteristics of returns data like negative skewness and leptokurtosis. Models with a time varying conditional volatility structure are often used, but the usual choice of error distributions may not consider the excess kurtosis present. We need to add more methods to the available toolkit for VaR estimation in order to resolve these issues. Since quantile behavior can be matched with

the standard deviation behavior, if the volatility modeling is based on a flexible structure, consequently, the flexibility will lend itself to quantile model behavior as well. Hence, a quantile regression based on the SV framework can lead us to a good model that captures the risk behavior effectively and efficiently.

The difficulty in the estimation of the SV model arises due to the addition of an innovation process. The exact likelihood cannot be solved analytically. The recursive model for the conditional VaR is intrinsically related to this issue. The main goal in this chapter is, therefore, to derive a quantile specification that approximates the evolution of the tail quantiles to be used by the RQ objective criterion directly without the likelihood computation in SV framework. This nonparametric approach is desirable, since no distributional assumptions need to be made. We propose linear optimal filters named SVLIN, A-SVLIN, to be discussed in Section 2.6.

Also, a direct application to VaR computation would be to use the nonlinear filtering methodology, introduced by Watanabe (1999), Fridman and Harris (1998) which produces the one step ahead conditional return density as a part of their algorithm to calculate the likelihood. One of the key aspects of this approach is that if we are able to get the conditional distribution of the returns itself, we can directly get the conditional quantiles. Calculating the quantiles from this distribution directly gives us a benchmark method to obtain the VaRs which can be further compared with what we have obtained using the linear filter.

2.3 Conditional Autoregressive Value at Risk class of models (CAViaR)

Let $\{y_t\}_{t=1}^{T}$ be the observed vector of portfolio returns and ξ_t a vector of time t observable variables. Let τ be the probability associated with VaR and β_τ be a vector of unknown parameters. Finally, let $q_t(\beta) \equiv q(\xi_{t-1}, \beta_\tau)$ denote the time t, τ^{th} quantile of the distribution of portfolio returns formed at time $t-1$, where the τ subscript is suppressed for notational convenience. Then, a general CAViaR specification is given by the following:

$$q_t(\beta) = \gamma_0 + \sum_{i=1}^{q} \gamma_i q_{t-i}(\beta) + \sum_{i=1}^{m} \alpha_i l(\xi_{t-i}, \varphi),$$

where $\beta = (\alpha, \gamma,\ \varphi)$ and l is a function of a finite number of lagged values of observables. The role of $l(\xi_{t-i}, \varphi)$ is to link $q_t(\beta)$ to the observables that belong to the information set.

A natural choice for ξ_{t-1} is lagged returns.
Some examples of CAViaR processes described are as follows:

Adaptive:

$$q_t(\beta_1) = q_{t-1}(\beta_1) + \beta_1\{[1 + \exp(G[y_{t-1} - q_{t-1}(\beta_1)])]^{-1} - \tau\}$$

Symmetric Absolute Value:

$$q_t(\beta) = \beta_1 + \beta_2 q_{t-1}(\beta) + \beta_3|y_{t-1}|$$

Asymmetric Slope:

$$q_t(\beta) = \beta_1 + \beta_2 q_{t-1}(\beta) + \beta_3(y_{t-1})^+ + \beta_4(y_{t-1})^-$$

Indirect GARCH:

$$q_t(\beta) = (\beta_1 + \beta_2 q_{t-1}^2(\beta) + \beta_3 y_{t-1}^2)^{\frac{1}{2}}$$

In the adaptive method, G is some positive finite number and the model is a smoothed version of a step function. As $G \to \infty$ the last term converges almost surely to $\beta_1[\mathrm{I}(\mathrm{y_{t-1}} \leq \mathrm{q_{t-1}}(\beta_1)) - \tau]$, where I(.) represents the indicator function. This model applies the following rule: whenever VaR is exceeded it should be immediately increased. When it is not exceeded one should decrease it, but very slightly. It increases the VaR by the same amount regardless of whether the returns exceed the VaR by a small or large margin.
The second and the fourth model respond symmetrically to past returns while the third allows different responses to positive and negative returns. All of the last three are mean reverting in the sense that the coefficient on the lagged VaR is not constrained to be one.
The indirect GARCH (iGARCH) model is the correctly specified model if the underlying data were truly a GARCH(1,1) with an i.i.d. error distribution. The Symmetric Absolute Value and Asymmetric Slope quantile models would be correctly specified by a GARCH process in which the standard deviation is modeled either symmetrically or asymmetrically with i.i.d. errors. For further details on the motivation behind these models, see Taylor (1986), Schwert (1989) and Engle(2002). The CAViaR specification is however more general than these GARCH models in the sense that non-i.i.d. error distributions can also be modeled.

2.4 Estimation of parameters by Regression Quantiles

The parameters of CAViaR models are estimated by Regression Quantiles objective criterion, as introduced by Koenker and Bassett (1978). Consider a sample of observations $y_1, y_2, \ldots, y_T$ generated by the following model:

$$y_t = x_t'\beta^0 + \varepsilon_{\tau t}$$

with $Quant_\tau(\varepsilon_{\tau t}|\mathbf{x_t}) = \mathbf{0}$, where $\mathbf{x_t}$ is a m-vector of regressors and $Quant_\tau(\varepsilon_{\tau t}|\mathbf{x_t})$ is the τ-quantile of $\varepsilon_{\tau t}$ conditional on $\mathbf{x_t}$. Let $q_t(\beta) \equiv x_t'\beta$. The τ^{th} regression quantile estimate is therefore $x_t'\hat{\beta}$. The parameter estimates are obtained by minimizing the Regression Quantile (RQ) objective function also known as the check function:

$$\min_\beta \frac{1}{T}\left[-\sum_{t=1}^{T}\{I(y_t < q_t(\beta)) - \tau\}(y_t - q_t(\beta))\right]$$

As a special case, regression quantiles include the least absolute deviation (LAD) model which is known to be more robust than OLS estimators whenever the errors have a fat tailed distribution.

2.5 Quantile Model Evaluation

2.5.1 Dynamic Quantile Test

The Dynamic Quantile (DQ) test can be used for evaluating the overall goodness of fit test for the estimated CAViaR process. An ideal condition of a VaR estimate is to create a sequence of i.i.d. indicator functions $I(y_t < q_t(\beta))$ from a possibly serially correlated heteroscedastic time series. This can lead to the testing based on whether the unconditional probabilities are correct and serially uncorrelated. However to account for the dependence, DQ tests consider the $\mathrm{Hit_t}(\beta) = \mathrm{I(y_t < q_t(\beta))} - \theta$, where the conditional expectation given any information till time, $t-1$ is zero. $\mathrm{Hit_t}(\beta)$ must be also uncorrelated with its own lagged values and with $q_t(\beta)$. The test takes into consideration any of the past information that affects the quantile estimates. Let T denote the in-sample data and N denote the out of sample data. The DQ out-of-sample test statistic is shown below. As $T \to \infty$, $N \to \infty$

$$\mathrm{DQ_O} = \frac{\mathrm{Hit_N}'(\hat{\beta}_\mathrm{T})\mathrm{X_N}(\hat{\beta}_\mathrm{T})[\mathrm{X_N}'(\hat{\beta}_\mathrm{T})\mathrm{X_N}(\hat{\beta}_\mathrm{T})]^{-1}\mathrm{X_N}'(\hat{\beta}_\mathrm{T})\mathrm{Hit_N}(\hat{\beta}_\mathrm{T})}{\theta(1-\theta)} \xrightarrow{\mathrm{d}} {\chi_\mathrm{p}}^2$$

where $\mathrm{X_n}(\hat{\beta}), \mathrm{n} = \mathrm{T}+1, \ldots, \mathrm{T}+\mathrm{N}$ is a typical row of $\mathrm{X_N}(\hat{\beta})$ with p columns, which is a p-vector measurable-Ω_n, and $\mathrm{Hit_N}(\hat{\beta}) = \left[\mathrm{Hit_{T+1}}(\hat{\beta}), \ldots, \mathrm{Hit_{T+N}}(\hat{\beta})\right]'$. The choice for X_n are the past lagged Hits, past VaRs.

The DQ in sample test statistic is given by:

$$\mathrm{DQ_I} = \frac{\mathrm{Hit_T}'(\hat{\beta})\mathrm{X_T}(\hat{\beta})[\mathrm{T}^{-1}\hat{\mathrm{M}}_\mathrm{T}\hat{\mathrm{M}}_\mathrm{T}']^{-1}\mathrm{X_T}'(\hat{\beta})\mathrm{Hit_T}(\hat{\beta})}{\mathrm{T}\theta(1-\theta)}$$

Let $\hat{M}_T = \mathrm{X_T}'(\hat{\beta}) - \left[(2\mathrm{T}\hat{\mathrm{c}}_\mathrm{T})^{-1}\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{I}\left(\left|\mathrm{y_t} - \mathrm{q_t}(\hat{\beta})\right| < \hat{\mathrm{c}}_\mathrm{T}\right)\mathrm{X_t}'(\hat{\beta})\nabla \mathrm{q_t}(\hat{\beta})\right]\hat{\mathrm{D}}_\mathrm{T}^{-1}\nabla' \mathrm{q}(\hat{\beta})$ where $\mathrm{X_t}(\hat{\beta})$, $t = 1, 2, \ldots, T$ is a typical row of $\mathrm{X_T}(\hat{\beta})$ is an m-vector measurable Ω_t and $\mathrm{Hit_T}(\hat{\beta}) = \left[\mathrm{Hit_1}(\hat{\beta}), \ldots, \mathrm{Hit_T}(\hat{\beta})\right]'$. The test statistic follows asymptotically a χ^2_m where m is the rank of the $\mathrm{X_T}(\hat{\beta})$ where $\hat{c}_T$ is the bandwidth of the k-nearest neighbour method.

2.5.2 Backtesting methods

Several backtesting methodologies are prevalent in the extant literature by matching the VaR forecasts with the portfolio returns. Based on these exceedences, several VaR validation tests can be made. Unconditional coverage test (UC) (Kupiec, 1995), Markov test (Christoffersen, 1998), conditional coverage test (CC) (Christoffersen, 1998) are used for evaluating the VaR models along with DQ tests in the data analysis, given later. The total number of exceedences, defined by $X = \sum_{t=1}^{T} I(y_t < -\mathrm{VaR_t})$ follows a Binomial(T,τ) under the null hypothesis when the exceedences are believed to be i.i.d. This idea forms the basis for the UC test. Markov test is an independence test to check for first order Markov dependence eg. clustering effect in the series. We set the indicator to 0 if VaR is not exceeded and to 1 otherwise. Let π_i be the probability of observing an exception conditional on state i the previous day. Let T_{ij} denote the number of days in which state j occured in a day with i on the previous day. Hence, to test an independence of exception on a day followed by another on the next day, the null hypothesis is $\pi_0 = \pi_1 = \pi = (T_{01} + T_{11})/T$. The test statistic is :

$$\mathrm{LR_{ind}} = -2\log[(1-\pi)^{(\mathrm{T_{00}}+\mathrm{T_{10}})}\pi^{\mathrm{T_{01}}+\mathrm{T_{11}}}] + 2\log[(1-\pi_0)^{\mathrm{T_{00}}}\pi_0^{\mathrm{T_{01}}}(1-\pi_1)^{\mathrm{T_{10}}}\pi_1^{\mathrm{T_{11}}}]$$

Furthermore, the conditional coverage test is the combined test statistic of the unconditional and Markov tests given by:

$$\mathrm{LR_{cc}} = \mathrm{LR_{uc}} + \mathrm{LR_{ind}}$$

distributed as χ^2_2. The $\mathrm{LR_{uc}} = -2\log[\mathrm{p}^\mathrm{X}(1-\mathrm{p})^{\mathrm{T}-\mathrm{X}}] + 2\log[(\mathrm{X/T})^\mathrm{X}(1-(\mathrm{X/T}))^{\mathrm{T}-\mathrm{X}}]$ which asymptotically follows a chi squared with one degrees of freedom under the null hypothesis.

The Weibull test is based on durations of the violations (Christoffersen and Pelletier, 2004). The chief idea is that given that the VaR model is correctly specified with coverage rate τ, the occurence of an excessive number of short durations (period of turbulence) and an excessive number of very long durations (period of tranquility) should signal a warning against the chosen VaR model. Under the null hypothesis that the model is correctly specified, the duration should have no memory with a mean duration of $1/\tau$ days. It is shown to have more power than the Markov test for testing independence in the Historical simulation method. However, it has not been used in this chapter because the out-of-sample chosen was 500 which had lower power than UC and Markov in the coverage area considered in their study. The Weibull test is used in Chapter 4 where the out-of-sample sizes are larger.

2.6 Optimal linear filter

In this section, we use linear filter methods to derive recursive equations for the conditional quantiles of SV processes.

2.6.1 Symmetric SV framework

Let y_t be the stochastic process of returns

$$y_t = \sigma_t \epsilon_t \qquad \epsilon_t \sim N(0,1) \tag{2.1}$$

where σ_t^2 is the volatility of y_t. In the simplest SV model framework, the log of volatility is expressed as an AR(1) model:

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + v_t \quad v_t \sim N(0, \sigma_v^2) \tag{2.2}$$

where ϵ_t and v_t are assumed to be independent of each other. We can write the linear state space form of the model given by (2.1) and (2.2) by taking logarithm of squared variables, as shown below:

$$\ln y_t^2 = \ln \sigma_t^2 + \ln \epsilon_t^2 \qquad t = 1, 2, \ldots, T \tag{2.3}$$

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + v_t \qquad v_t \sim N(0, \sigma_v^2) \tag{2.4}$$

The $\ln(y_t^2)$ process is an AR(1) process with added white noise. Arranging the above equations, the $\ln y_t^2$ process can be represented as an ARMA(1,1). Under the current setup, an immediate goal is to seek a CAViaR representation of conditional quantiles with a best

fit, when the underlying data is generated from a SV model. This goal can be achieved by finding the best linear predictor of the $\ln \sigma_t^2$ process conditioned on the past returns by minimizing the expected mean squared error, which would eventually suggest the quantile regression form. For the state space representation considered above, the Kalman filter yields the best linear predictor (ie., the best predictor with the lowest MSE among the class of linear estimators (Anderson and Moore, 1979)). The Kalman filter steps are given in Section 4.1 of Chapter 4. Based on (2.3) and (2.4), and following the same notation as in Chapter 4, let $x_t = \ln(\sigma_t^2)$ and $\eta_t = \ln(\epsilon_t^2)$. η_t has a finite mean and variance denoted by μ_η and σ_η^2 respectively. When ϵ_t is Gaussian, $\mu_\eta = -1.27$ and $\sigma_\eta^2 = \frac{\pi^2}{2}$. Let $x_{t|t-1}$ and $P_{t|t-1}$ denote respectively the optimal linear estimator and the variance of x_t given information till time $t-1$, and $x_{t|t}$ and $P_{t|t}$ denote respectively the updated optimal linear estimator and the variance of x_t given information till time t. Combining the one step ahead prediction and updating equations into a single recursion step of the Kalman filter, produces,

$$x_{t+1|t} = \alpha_0 + \alpha_1 \left[x_{t|t-1} + \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\eta^2} \left(\ln y_t^2 - x_{t|t-1} - \mu_\eta \right) \right]$$

and

$$P_{t+1|t} = \alpha_1^2 P_{t|t-1} \left(1 - \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\eta^2} \right) + \sigma_v^2.$$

For stationarity, $P_{t+1|t} = P_{t|t-1} = \sigma^2$, say, the second equation can be solved for σ^2, and with the result plugged into the first equation:

$$\ln \widehat{(\sigma_{t+1}^2)}_{\text{lin}} = \eta_1 + \eta_2 \ln(\sigma_t^2) + \eta_3 \ln y_t^2$$

This suggests a linear quantile regression form as follows:

$$\ln q_t^2(\beta) = \beta_1 + \beta_2 \ln q_{t-1}^2(\beta) + \beta_3 \ln y_{t-1}^2$$

We call this the SVLIN model.

2.6.2 Asymmetric SV framework

There are two different Asymmetric SV models in the literature, one arising as the discretized version of the continuous time model (Harvey and Shephard, 1996), and another slightly modified version introduced by Jacquier *et al.* (2004). We call the latter ASV2 model. Yu (2005) argued that the ASV model (Harvey and Shephard, 1996) given by (2.5, 2.6) produced better fits to data empirically using MCMC techniques. The competing

model (ASV2) introduced by Jacquier *et al.* (2004), with ASV has intertemporal instead of contemporaneous correlation. This makes the ASV model generate a Martingale Difference sequence.

$$y_t = \sigma_t \epsilon_t \tag{2.5}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \sigma_{t-1}^2 + v_t \tag{2.6}$$

where ϵ_t and v_t are i.i.d. N(0,1) and $\text{Corr}(\epsilon_t, v_{t+1}) = \rho$. For ASV2 model, the only change in assumption is $\text{Corr}(\epsilon_t, v_t) = \rho$. Further, Yu (2005) shows with a nonlinear state space transformation that the ASV gives a clear elicitation of the negative correlation of returns with volatility, i.e, the fall in the stock price leads to an increase in the volatility. Hence, ASV model is chosen over ASV2 as a basis for a quantile specification that takes into account the asymmetric effect. If the joint distribution of the error terms are symmetric, then the log transformation used to obtain the measurement equation results in the loss of information. However, Harvey and Shephard (1996) show that by retaining the sign of the returns along with the absolute values gives the same likelihood function,

$$f(\mathbf{y_T}|\theta) = \prod_{t=1}^{T} f(y_t|s_t, \mathbf{y_{T-1}}, \theta) f(s_t|\mathbf{y_{T-1}}, \theta)$$

where $\mathbf{y_T} = \{y_1, y_2, \ldots, y_T\}$, s_t is the sign function assigned with each return being positive or negative. Since the ϵ_t's are symmetrically distributed, $f(s_t|\mathbf{y_{T-1}}, \theta) = 0.5$, this leads to maximizing the conditional density function given by

$$\prod_{t=1}^{T} f(|y_t| \;\Big|\; s_t, \mathbf{y_{T-1}}, \theta)$$

Hence, the Kalman filter applied to the ASV model is used to find the filtering equations. Following Harvey and Shephard (1996), based on the following transformations: $h_t = \ln \sigma_t^2 - \mu_h$, $\mu_h = \alpha_0/(1-\alpha_1)$, $w = \mu_h + E(\ln \epsilon_t^2)$, $\xi_t = \ln \epsilon_t^2 - E(\ln \epsilon_t^2)$, $\mu^* = E(v_{t+1}|\text{sgn}(\epsilon_t) = +) = -\text{E}(\text{v}_{\text{t+1}}|\text{sgn}(\epsilon_\text{t}) = -)$, $\gamma^* = \text{Cov}(\text{v}_{\text{t+1}}, \xi_\text{t}|\text{sgn}(\epsilon_\text{t}) = +) = -\text{Cov}(\text{v}_{\text{t+1}}, \xi_\text{t}|\text{sgn}(\epsilon_\text{t}) = -)$, $\text{Var}(\text{v}_{\text{t+1}}) = \text{E}(\text{v}_{\text{t+1}}^2|\text{sgn}(\epsilon_\text{t}) = +) - (\text{E}(\text{v}_{\text{t+1}}|\text{sgn}(\epsilon_\text{t}) = +)^2) = \sigma_\text{v}^2 - \mu^{*2}$, the measurement and state equations are:

$$\log y_t^2 = w + h_t + \xi_t, h_{t+1} = \left(\alpha_1 - \frac{\gamma^* s_t}{\sigma_\xi^2}\right) h_t + s_t \left\{\mu^* + \frac{\gamma^*}{\sigma_\xi^2}\left(\log y_t^2 - w\right)\right\} + v_{t+1}^+$$

Denoting the linear optimal predictor of h_t as $h_{t|t-1}$ and its variance as $P^h_{t|t-1}$, and applying the Kalman filter:

$$\left(\begin{array}{c} \xi_t \\ v^+_{t+1} \end{array} \right) \Bigg| \, s_t \sim N \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma^2_\xi & 0 \\ 0 & \sigma^2_v - \mu^{*2} - \dfrac{\gamma^{*2}}{\sigma^2_\xi} \end{array} \right) \right)$$

$$\left(\begin{array}{c} h_t \\ \log y^2_t \end{array} \right) \Bigg| \log \mathbf{y^2_{t-1}}, \mathbf{s_t} \sim N \left(\left(\begin{array}{c} h_{t|t-1} \\ h_{t|t-1} + w \end{array} \right), \left(\begin{array}{cc} P^h_{t|t-1} & P^h_{t|t-1} \\ P^h_{t|t-1} & P^h_{t|t-1} + \sigma^2_\xi \end{array} \right) \right)$$

$$h_t | \log \mathbf{y^2_t}, s_t \sim N \left(h_{t|t-1} + \frac{P^h_{t|t-1}}{P^h_{t|t-1} + \sigma^2_\xi} \left(\log y^2_t - h_{t|t-1} - w \right), P^h_{t|t-1} \left(1 - \frac{P^h_{t|t-1}}{P^h_{t|t-1} + \sigma^2_\xi} \right) \right)$$

$$\begin{aligned} h_{t+1} | \log \mathbf{y^2_t}, s_t \quad \sim \quad & N \Bigg(s_t \left(\mu^* + \frac{\gamma^*}{\sigma^2_\xi} (\log y^2_t - w) \right) + \left(\alpha_1 - \frac{\gamma^* s_t}{\sigma^2_\xi} \right) \times \\ & \left[h_{t|t-1} + \frac{P^h_{t|t-1}}{P^h_{t|t-1} + \sigma^2_\xi} \left(\log y^2_t - h_{t|t-1} - w \right) \right], \\ & \left(\alpha_1 - \frac{\gamma^* s_t}{\sigma^2_\xi} \right)^2 P^h_{t|t-1} \left(1 - \frac{P^h_{t|t-1}}{P^h_{t|t-1} + \sigma^2_\xi} \right) + \sigma^2_v - \mu^{*2} - \frac{\gamma^{*2}}{\sigma^2_\xi} \Bigg) \end{aligned}$$

Hence, to obtain the optimal linear filter, the recursion step used is as follows:

$$h_{t+1|t} = s_t \mu^* + \alpha_1 h_{t|t-1} + \frac{s_t \gamma^* + \alpha_1 P^h_{t|t-1}}{P^h_{t|t-1} + \sigma^2_\xi} \left(\log y^2_t - h_{t|t-1} - w \right) \tag{2.7}$$

With the stationarity assumption, (2.7) suggests a linear quantile regression form as follows:

$$\ln q^2_t(\beta) = \beta_1 + \beta_2 \ln q^2_{t-1}(\beta) + \beta_3 \ln y^2_{t-1} + \beta_4 s_{t-1} \ln y^2_{t-1} + \beta_5 s_{t-1} \ln q^2_{t-1}(\beta)$$

We call this the A-SVLIN model. The proposed models along with the CAViaR is used to compute the VaRs under SV. The VaR performance from each of these models are compared with the quantiles obtained from the non linear filtering (NF) scheme. The NF scheme uses the conditional densities in its filter, hence, the conditional VaR's can be obtained directly.

2.7 NF as Benchmark Method

The non linear filtering (NF) method developed by Watanabe (1999), Fridman and Harris (1998) is a filtering scheme to evaluate the exact likelihood. The filtering technique

is based on the Bayes theorem. Let us define $x_t = \ln(\sigma_t^2)$ and $\mathbf{y_t} = (y_1, y_2, \ldots y_t)$. Then from equations (2.1) and (2.2), we have

$$p(x_t|x_{t-1}, \mathbf{y_{t-1}}) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left[-\frac{(x_t - \alpha_0 - \alpha_1 x_{t-1})^2}{2\sigma_v^2}\right] \tag{2.8}$$

$$p(y_t|x_t, \mathbf{y_{t-1}}) = \frac{1}{\sqrt{2\pi \exp(x_t)}} \exp\left[-\frac{(y_t^2)}{2\exp(x_t)}\right] \tag{2.9}$$

Although the log-SV-AR1 model used in this context allows the following simplifications: $p(x_t|x_{t-1}, \mathbf{y_{t-1}}) = p(x_t|x_{t-1})$ and $p(y_t|x_t, \mathbf{y_{t-1}}) = p(y_t|x_t)$, it might be noted that this algorithm provides a setup for easily adding several extensions to SV models since no transformation of the model is required. Once a judicious choice of the error distributions is made, the following algorithm becomes ready for analysis. Applying Bayes theorem, the filter is given by:

One-step ahead prediction

$$p(x_t|\mathbf{y_{t-1}}) = \int_{-\infty}^{\infty} p(x_t, x_{t-1}|\mathbf{y_{t-1}}) dx_{t-1} \tag{2.10}$$

$$= \int_{-\infty}^{\infty} p(x_t|x_{t-1}\mathbf{y_{t-1}}) p(x_{t-1}|\mathbf{y_{t-1}}) dx_{t-1} \tag{2.11}$$

Updating

$$p(x_t|\mathbf{y_t}) = p(x_t|y_t, \mathbf{y_{t-1}}) = \frac{p(x_t, y_t|\mathbf{y_{t-1}})}{p(y_t|\mathbf{y_{t-1}})} = \frac{p(y_t|x_t, \mathbf{y_{t-1}}) p(x_t|\mathbf{y_{t-1}})}{p(y_t|\mathbf{y_{t-1}})} \tag{2.12}$$

where the denominator is given by

$$p(y_t|\mathbf{y_{t-1}}) = \int_{-\infty}^{\infty} p(y_t, x_t|\mathbf{y_{t-1}}) dx_t \tag{2.13}$$

$$= \int_{-\infty}^{\infty} p(y_t|x_t, \mathbf{y_{t-1}}) p(x_t|\mathbf{y_{t-1}}) dx_t \tag{2.14}$$

Although (2.8) and (2.9) are known, the above integrals cannot be solved analytically. As suggested by Kitagawa, the densities $p(x_t|\mathbf{y_{t-1}})$ and $p(x_t|\mathbf{y_t})$ for $t = 1, 2, \ldots T$ are approximated by piecewise linear functions by choosing the number of segments, location and number of node points and the density values at these points. After the $N + 1$ nodes were selected for each time point t, the node points, sorted in order of size are denoted by

$x_t^{(0)}, x_t^{(1)}, \ldots, x_t^{(N)}$ for each of time point $t = 1, 2, \ldots, T$. Using the trapezoidal rule, these integrals are approximated in the following manner: One step ahead prediction:

$$\begin{aligned} p(x_t^{(i)}|\mathbf{y_{t-1}}) &= \int_{-\infty}^{\infty} p(x_t^{(i)}|x_{t-1}, \mathbf{y_{t-1}}) p(x_{t-1}|\mathbf{y_{t-1}}) dx_{t-1} \qquad (2.15) \\ &\approx \sum_{n=1}^{N} \int_{x_{t-1}^{(n-1)}}^{x_{t-1}^{(n)}} p(x_t^{(i)}|x_{t-1}, \mathbf{y_{t-1}}) p(x_{t-1}|\mathbf{y_{t-1}}) dx_{t-1} \\ &\approx \frac{1}{2} \sum_{n=1}^{N} (x_{t-1}^{(n)} - x_{t-1}^{(n-1)}) [p(x_t^{(i)}|x_{t-1}^{(n-1)}, \mathbf{y_{t-1}}) p(x_{t-1}^{(n-1)}|\mathbf{y_{t-1}}) + \\ &\quad p(x_t^{(i)}|x_{t-1}^{(n)}, \mathbf{y_{t-1}}) p(x_{t-1}^{(n)}|\mathbf{y_{t-1}})] \quad i = 0, 1, \ldots, N \end{aligned}$$

Updating:

$$p(x_t^{(i)}|\mathbf{y_t}) = \frac{p(y_t|x_t^{(i)}, \mathbf{y_{t-1}}) p(x_t^{(i)}|\mathbf{y_{t-1}})}{p(y_t|\mathbf{y_{t-1}})} \quad i = 0, 1, \ldots, N \qquad (2.16)$$

where the denominator can be approximated similarly. Using the knowledge that the marginal distribution of $p(x_t)$ follows a $N(\alpha_0/(1-\alpha_1), \sigma_v^2/(1-\alpha_1^2))$, we can approximate the initial starting point of the recursion by letting $p(x_1|0) \equiv p(x_1)$. We obtain the $p(y_t|\mathbf{y_{t-1}})$ as a byproduct of this recursion, hence we can evaluate the corresponding conditional quantiles for each period t. VaR at time t with probability level α $(0 < \alpha < 1)$ is defined as:

$$\mathrm{VaR_t}(\alpha) = -\inf\{\mathrm{y_t} : \mathrm{F_t}(\mathrm{y_t}|\mathbf{y_{t-1}}) \geq \alpha\},$$

where $F(.)$ denotes the c.d.f. of $p(y_t|\mathbf{y_{t-1}})$. Also, they can be used to obtain the exact likelihood:

$$\begin{aligned} \ln L &= \ln[p(\mathbf{y_T})] \qquad (2.17) \\ &= \sum_{t=1}^{T} \ln(p(y_t|\mathbf{y_{t-1}})) \qquad (2.18) \end{aligned}$$

Once the error due to the linear approximations are made negligible, the NFML (nonlinear filtering maximum likelihood) method yields the exact likelihood. A crucial step is to obtain the number and location of nodes in such a way so that it can perform well and be computationally viable, without compromising the results. Watanabe (1999) shows that with $N = 50$, it is possible to obtain good results when the node locations are chosen judiciously. Hence, this method serves the dual purpose of obtaining likelihood based estimates and also quantile estimates. For VaR calculations, we implement this method as a benchmark to obtain the conditional VaR's. We gauge the performance of these methods by data analysis.

2.8 Empirical Study

An empirical study based on daily prices demonstrate the performance of the SVLIN model in comparison to the existing CAViaR models.

Six daily closing stock prices comprising Merck (MRK), Merrill Lynch & Co. (MER), Medtronic (MDT), General Electric (GE), Ford (F), Genentech (DNA) are obtained from Yahoo! Finance from 7/86 to 6/08, and are chosen to test the performance of the quantile models. After discarding missing values and computing the log returns, the total number of observations are 3968, which are further split into 2968 in-sample data and 1000 out-of-sample data. Figure 2.1 shows the log returns of the six stocks. The graphs show the phenomenon of volatility clustering. The quantile estimates from each

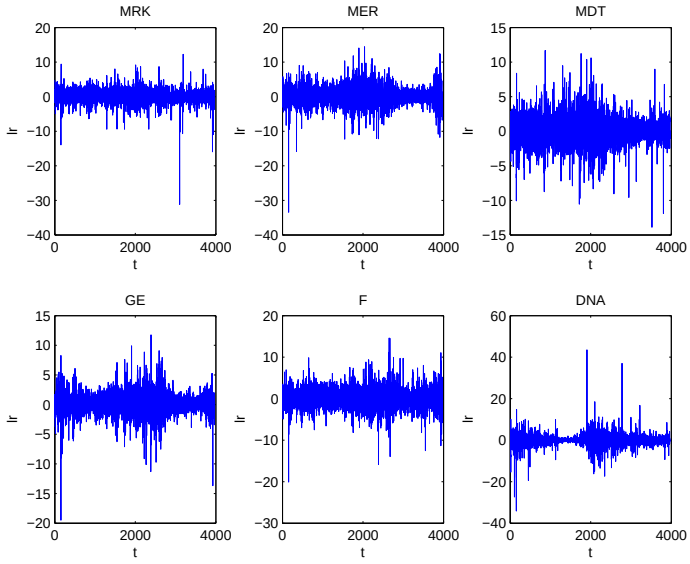

Figure 2.1: *Log returns of 6 stocks*

of the quantile model specifications are obtained by minimizing the Regression Quantile (RQ) objective criterion. The initialization of $q_1(\beta)$ is obtained by finding the empirical τ^{th} quantile of the in-sample data. The initial starting parameter values are obtained by generating 10,000 Uniform(0,1) random vectors of the parameter length. We compute the RQ objective function value for each of these vectors and choose 5 vectors with the lowest RQ criterion as the starting values for the optimization. We use fminsearch function as the optimization tool in Matlab. The covariates of the Dynamic Quantile (DQ, see section 2.5) out-of-sample test are a constant, a VaR forecast and the first four lagged hits. Following

Engle and Manganelli (2004), the covariance matrices $\hat{D}_T$ and $\hat{M}_T$ are computed using k-nearest neighbor estimators, with $k = 40$ for 1% and $k = 60$ for 5% VaR cases. In Tables 2.1 – 2.12 the estimated parameters, the standard errors, their p-values, the minimum RQ criterion, percentages of hits, both in-sample and out-of-sample, and the DQ test results are reported.

Table 2.1: *Comparison of CAViaR, SVLIN models for MRK at 1% VaR. Significant coefficients at 5% are formatted in bold.*

	MRK				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.2285	0.2096	**0.8832**	**0.2281**	**0.2916**
Std. Err.	0.2044	0.1617	0.5207	0.1246	0.1141
P-Values	0.1319	0.0974	0.0449	0.0336	0.0053
β_2	**0.9246**	**0.9397**	**0.9403**	**0.9267**	**0.9045**
Std. Err.	0.0564	0.0481	0.0197	0.0399	0.0357
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	0.0968	-0.0028	0.1318	**0.0349**	**0.0448**
Std. Err.	0.0909	0.0707	0.4020	0.0196	0.0166
P-Values	0.1433	0.4840	0.3715	0.0376	0.0035
β_4		0.1140			**-0.0427**
Std. Err.		0.0906			0.0210
P-Values		0.1041			0.0209
β_5					0.0092
Std. Err.					0.0143
P-Values					0.2589
RQ	193.1013	191.1910	193.3665	193.1425	189.9878
Hits in(%)	0.9771	0.9771	0.9771	0.9771	0.9771
Hits out(%)	0.8000	0.8	0.9	1.3	1.3
DQ in sample (pval)	0.6891	0.6834	0.3775	0.4648	0.7740
DQ out-of-sample (pval)	0.9611	0.9864	0.9820	0.2890	0.2180

For all the models, β_2, the autoregressive parameter is always significant. This confirms that the clustering phenomenon is relevant also at the tails. The percentage of in-sample hits for all the models in all six data scenarios demonstrate the precision and applicability of all these models in terms of fitting. The RQ objective values for the different models are comparable indicating that the SVLIN model can be used along with CAViaR models for analysis. An interesting feature is that the SVLIN and A-SVLIN models have significant coefficients at the 5% level, in most of the cases. All the models seem to fit

Table 2.2: *Comparison of CAViaR, SVLIN models for MRK at 5% VaR. Significant coefficients at 5% are formatted in bold.*

	MRK				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.0463	**0.0896**	0.1740	**0.1319**	**0.1298**
Std. Err.	0.0410	0.0414	0.1460	0.0533	0.0640
P-Values	0.1293	0.0152	0.1168	0.0067	0.0212
β_2	**0.9485**	**0.9324**	**0.9466**	**0.9364**	**0.9372**
Std. Err.	0.0235	0.0242	0.0181	0.0248	0.0297
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.0674**	0.0071	0.0740	**0.0397**	**0.0368**
Std. Err.	0.0266	0.0309	0.2513	0.0130	0.0191
P-Values	0.0056	0.4094	0.3842	0.0011	0.0269
β_4		**0.1364**			**-0.0230**
Std. Err.		0.0394			0.0092
P-Values		0.0003			0.0062
β_5					-0.0152
Std. Err.					0.0097
P-Values					0.0582
RQ	618.2274	610.8398	620.389	619.6961	614.9603
Hits in(%)	5.0202	5.0539	5.0202	5.0202	5.0202
Hits out(%)	4.6	3.5	3	5.4	4.8
DQ in sample (pval)	0.0015	0.0191	0.0073	0.0001	0.0046
DQ out-of-sample (pval)	0.0626	0.1329	0.0208	0.0086	0.0569

Table 2.3: *Comparison of CAViaR, SVLIN models for MER at 1% VaR. Significant coefficients at 5% are formatted in bold.*

	MER				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.5968	0.4126	**0.9607**	**0.8127**	**0.4831**
Std. Err.	0.5068	0.4136	0.1243	0.1958	0.1727
P-Values	0.1195	0.1593	0.0042	0.0000	0.0026
β_2	**0.7891**	**0.8653**	**0.8344**	**0.7584**	**0.8528**
Std. Err.	0.1275	0.0929	0.0289	0.0534	0.0512
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.3789**	0.0267	0.5997	**0.1244**	**0.0659**
Std. Err.	0.1692	0.0664	0.9621	0.0203	0.0267
P-Values	0.0126	0.3436	0.2665	0.0000	0.0068
β_4		**0.3953**			**-0.0633**
Std. Err.		0.1473			0.0220
P-Values		0.0036			0.0020
β_5					-0.0157
Std. Err.					0.0215
P-Values					0.2324
RQ	269.9123	262.2050	272.3408	269.0879	260.4925
Hits in(%)	1.0108	1.0445	0.9771	0.9771	0.9434
Hits out(%)	0.4	0.4	0.4	0.6	0.6
DQ in sample (pval)	0.0545	0.6902	0.0001	0.0327	0.7363
DQ out-of-sample (pval)	0.4590	0.5285	0.5793	0.7803	0.7366

Table 2.4: *Comparison of CAViaR, SVLIN models for MER at 5% VaR. Significant coefficients at 5% are formatted in bold.*

	MER				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	**0.1066**	**0.1588**	**0.4889**	**0.2802**	**0.2944**
Std. Err.	0.0547	0.0711	0.2064	0.0574	0.0405
P-Values	0.0256	0.0128	0.0089	0.0000	0.0000
β_2	**0.8947**	**0.8858**	**0.9072**	**0.8749**	**0.8756**
Std. Err.	0.0304	0.0278	0.0140	0.0229	0.0152
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.1469**	**0.0769**	0.1258	**0.0836**	**0.0589**
Std. Err.	0.0432	0.0380	0.1016	0.0189	0.0105
P-Values	0.0003	0.0215	0.1079	0.0000	0.0000
β_4		**0.2075**			**-0.0433**
Std. Err.		0.0519			0.0116
P-Values		0.0000			0.0001
β_5					**-0.0196**
Std. Err.					0.0073
P-Values					0.0034
RQ	839.4421	830.5276	841.9084	849.0707	831.1015
Hits in(%)	4.9528	5.0539	5.0202	5.0202	4.9865
Hits out(%)	4.7	3.6	3.9	5.3	4.3
DQ in sample (pval)	0.9011	0.9981	0.5082	0.0593	0.808
DQ out-of-sample (pval)	0.0085	0.0008	0.0007	0.0043	0.0035

Table 2.5: *Comparison of CAViaR, SVLIN models for MDT at 1% VaR. Significant coefficients at 5% are formatted in bold.*

	MDT				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.5881	0.2831	**3.8619**	**0.6199**	0.5304
Std. Err.	0.4090	0.2356	2.3091	0.3373	0.3912
P-Values	0.0752	0.1148	0.0472	0.0330	0.0876
β_2	**0.7374**	**0.8656**	**0.6804**	**0.8029**	**0.8304**
Std. Err.	0.0960	0.0669	0.0849	0.1017	0.1209
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.4974**	0.1078	0.3203	**0.1036**	**0.0887**
Std. Err.	0.1146	0.1361	1.0705	0.0335	0.0417
P-Values	0.0000	0.2140	0.1087	0.0010	0.0166
β_4		**0.4017**			-0.0167
Std. Err.		0.0917			0.0259
P-Values		0.0000			0.2601
β_5					**-0.0302**
Std. Err.					0.0110
P-Values					0.0031
RQ	203.0630	200.1670	206.9602	202.5136	197.9110
Hits in(%)	0.9771	0.9771	0.9771	1.0108	1.0108
Hits out(%)	0.5	0.5	0.4	0.5	0.5
DQ in sample (pval)	0.4664	0.4232	0.0137	0.0251	0.4752
DQ out-of-sample (pval)	0.7819	0.7402	0.6996	0.8230	0.8130

Table 2.6: *Comparison of CAViaR, SVLIN models for MDT at 5% VaR. Significant coefficients at 5% are formatted in bold.*

	MDT				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	**0.0763**	**0.0489**	**0.1551**	**0.2365**	**0.19**
Std. Err.	0.0367	0.0290	0.0826	0.0746	0.0635
P-Values	0.0189	0.0456	0.0301	0.0008	0.0014
β_2	**0.9153**	**0.9339**	**0.9382**	**0.8883**	**0.9119**
Std. Err.	0.0157	0.0144	0.0097	0.0333	0.0292
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.1160**	**0.0599**	**0.1096**	**0.0669**	**0.0519**
Std. Err.	0.0222	0.0250	0.0656	0.0124	0.0176
P-Values	0.0000	0.0083	0.0474	0.0000	0.0016
β_4		**0.1390**			**-0.0225**
Std. Err.		0.0209			0.0122
P-Values		0.0000			0.0324
β_5					-0.0157
Std. Err.					0.0139
P-Values					0.1280
RQ	674.8661	671.7328	678.5569	679.2736	676.2068
Hits in(%)	5.0202	5.0202	5.0202	4.9865	5.0539
Hits out(%)	2.5	2.6	2.1	3	3.1
DQ in sample (pval)	0.6614	0.7874	0.6482	0.4348	0.6487
DQ out-of-sample (pval)	0.0236	0.0312	0.0035	0.0766	0.0786

Table 2.7: *Comparison of CAViaR, SVLIN models for GE at* 1% *VaR. Significant coefficients at 5% are formatted in bold.*

	GE				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.0981	**0.1307**	0.3857	**0.2458**	**0.2280**
Std. Err.	0.0663	0.0552	0.2904	0.0850	0.0548
P-Values	0.0696	0.0090	0.0921	0.0019	0.0000
β_2	**0.9170**	**0.9084**	**0.9015**	**0.9171**	**0.9232**
Std. Err.	0.0221	0.0193	0.0195	0.0277	0.0187
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.1828**	-0.0218	0.4656	**0.0638**	**0.0423**
Std. Err.	0.0613	0.0404	0.9616	0.0115	0.0175
P-Values	0.0014	0.2950	0.3141	0.0000	0.0077
β_4		**0.4379**			**-0.0567**
Std. Err.		0.1100			0.0193
P-Values		0.0000			0.0017
β_5					**-0.0368**
Std. Err.					0.0137
P-Values					0.0036
RQ	190.8196	179.6669	193.0525	191.9671	182.5715
Hits in(%)	1.0108	0.9771	1.0108	0.9771	0.9434
Hits out(%)	0.7	0.4	0.6	0.8	0.8
DQ in sample (pval)	0.0002	0.6778	0.0495	0.0001	0.0606
DQ out-of-sample (pval)	0.9843	0.6773	0.9396	0.9675	0.9287

Table 2.8: *Comparison of CAViaR, SVLIN models for GE at* 5% *VaR. Significant coefficients at 5% are formatted in bold.*

	GE				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.0238	**0.0516**	**0.0803**	**0.1383**	**0.2350**
Std. Err.	0.0148	0.0287	0.0437	0.0296	0.0603
P-Values	0.0536	0.0360	0.0331	0.0000	0.0000
β_2	**0.9172**	**0.9085**	**0.9101**	**0.9287**	**0.8832**
Std. Err.	0.0224	0.0174	0.0074	0.0155	0.0311
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.14**	0.0624	**0.1798**	**0.0651**	**0.0782**
Std. Err.	0.0424	0.0434	0.0744	0.0116	0.0262
P-Values	0.0005	0.0752	0.0079	0.0000	0.0014
β_4		**0.2214**			**-0.0486**
Std. Err.		0.0201			0.0172
P-Values		0.0000			0.0024
β_5					**-0.0393**
Std. Err.					0.0099
P-Values					0.0000
RQ	598.2705	591.4999	598.5543	606.0340	598.7831
Hits in(%)	4.9865	5.0202	5.0539	4.9865	4.9191
Hits out(%)	5.0000	3.5	4.0000	6.7	3.5
DQ in sample (pval)	0.0614	0.8737	0.0676	0.0012	0.8647
DQ out-of-sample (pval)	0.0037	0.0809	0.0500	0.0002	0.0027

Table 2.9: *Comparison of CAViaR, SVLIN models for F at 1% VaR. Significant coefficients at 5% are formatted in bold.*

	F				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.0147	0.0605	0.1178	**0.2271**	**0.1789**
Std. Err.	0.0715	0.0550	0.3450	0.0805	0.0887
P-Values	0.4183	0.1356	0.3664	0.0024	0.0219
β_2	**0.9355**	**0.9501**	**0.9217**	**0.9262**	**0.9430**
Std. Err.	0.0149	0.0115	0.0126	0.0248	0.0276
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.1881**	0.0228	0.4249	**0.0552**	**0.0356**
Std. Err.	0.0348	0.0398	0.3826	0.0175	0.0195
P-Values	0.0000	0.2832	0.1334	0.0008	0.0339
β_4		**0.2135**			-0.0221
Std. Err.		0.0534			0.0207
P-Values		0.0000			0.1427
β_5					-0.0138
Std. Err.					0.0115
P-Values					0.1146
RQ	214.5552	208.7544	215.7493	217.3162	213.0798
Hits in(%)	0.9771	0.9771	0.9434	1.0108	1.0108
Hits out(%)	1.4	1.2	1.4	1.0000	0.9
DQ in sample (pval)	0.6202	0.6655	0.9064	0.4587	0.5434
DQ out-of-sample (pval)	0.7276	0.6549	0.7434	0.9980	0.9589

Table 2.10: *Comparison of CAViaR, SVLIN models for F at 5% VaR. Significant coefficients at 5% are formatted in bold.*

	F				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	0.0935	**0.1231**	**0.4843**	**0.1172**	**0.1523**
Std. Err.	0.0763	0.0454	0.2928	0.0442	0.07
P-Values	0.1101	0.0034	0.0490	0.0040	0.0148
β_2	**0.8880**	**0.8770**	**0.8709**	**0.9448**	**0.9305**
Std. Err.	0.0452	0.0259	0.0248	0.0205	0.0313
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.1703**	**0.1074**	0.2276	**0.0472**	**0.0478**
Std. Err.	0.0533	0.0317	0.2342	0.0126	0.0156
P-Values	0.0007	0.0003	0.1656	0.0001	0.0011
β_4		**0.2452**			**-0.0259**
Std. Err.		0.0481			0.0129
P-Values		0.0000			0.0222
β_5					-0.0127
Std. Err.					0.0122
P-Values					0.1491
RQ	725.0152	718.0551	726.7447	732.3231	724.9911
Hits in(%)	5.0202	4.9865	4.9865	5.0202	4.9865
Hits out(%)	5.4	5.2	4.6	5.8	5
DQ in sample (pval)	0.3617	0.7415	0.2817	0.0306	0.9583
DQ out-of-sample (pval)	0.3959	0.4632	0.3093	0.0080	0.9783

Table 2.11: *Comparison of CAViaR, SVLIN models for DNA at* 1% *VaR. Significant coefficients at 5% are formatted in bold.*

	DNA				
1% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	-0.0174	0.0198	-0.0150	**0.1906**	**0.3685**
Std. Err.	0.0137	0.0339	0.0141	0.1159	0.0965
P-Values	0.1019	0.2796	0.1445	0.0500	0.0001
β_2	**0.9326**	**0.9075**	**0.9694**	**0.9387**	**0.8878**
Std. Err.	0.0103	0.0232	0.0045	0.0346	0.0311
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.2554**	**0.1258**	0.2075	**0.0649**	**0.1079**
Std. Err.	0.0354	0.0532	0.9811	0.0315	0.0237
P-Values	0.0000	0.0090	0.4162	0.0197	0.0000
β_4		**0.5358**			0.0041
Std. Err.		0.1174			0.0676
P-Values		0.0000			0.4758
β_5					-0.0419
Std. Err.					0.0358
P-Values					0.1208
RQ	301.0565	290.7673	310.1899	298.3042	283.6742
Hits in(%)	0.9771	0.9771	1.0445	0.9771	1.0108
Hits out(%)	1.3	1.3	1.3	1.7	1.5
DQ in sample (pval)	0.8793	0.9287	0.7774	0.5992	0.8749
DQ out-of-sample (pval)	0.0738	0.0729	0.2534	0.0000	0.0001

Table 2.12: *Comparison of CAViaR, SVLIN models for DNA at* 5% *VaR. Significant coefficients at 5% are formatted in bold.*

	DNA				
5% VaR	SAV	ASYMM	iGARCH	SVLIN	A-SVLIN
β_1	-0.0002	0.0048	0.0012	**0.1133**	**0.1488**
Std. Err.	0.0031	0.0076	0.0025	0.0323	0.0421
P-Values	0.4761	0.2645	0.3104	0.0002	0.0002
β_2	**0.9789**	**0.9546**	**0.9875**	**0.9459**	**0.93**
Std. Err.	0.0044	0.0124	0.0010	0.0152	0.0195
P-Values	0.0000	0.0000	0.0000	0.0000	0.0000
β_3	**0.0406**	**0.0610**	0.0215	**0.0525**	**0.0698**
Std. Err.	0.0069	0.0280	0.0418	0.0152	0.0166
P-Values	0.0000	0.0146	0.3038	0.0003	0.0000
β_4		**0.1106**			-0.0151
Std. Err.		0.0469			0.0309
P-Values		0.0092			0.3121
β_5					-0.0154
Std. Err.					0.0167
P-Values					0.1786
RQ	885.3730	884.4096	905.8684	881.2903	880.0991
Hits in(%)	5.0539	4.9865	5.1213	4.8854	4.9191
Hits out(%)	5.6	5.4	5.6	6.7	6.2
DQ in sample (pval)	0.0035	0.2349	0.0001	0.0378	0.1402
DQ out-of-sample (pval)	0.3214	0.2070	0.3717	0.0015	0.0002

the data well as measured by the Hit in sample performance and RQ objective value. It is important to note that the RQ value is usually the minimum in the Asymmetric-SVLIN cases, especially for the 1% VaR. That the ASYMM and A-SVLIN perform better than the others in the 5% case indicate that the asymmetric pattern is a nonnegligible feature in the evolution of quantiles at the tail region eg., F, MDT, DNA. The adaptive model usually performs the worst, and has not been reported here. For the 1% case, all the stocks are successfully fitted by at least one of the specifications. One exception is that the Ford data is fitted very well by all. The reason could be that the out of sample data continues to follow the same pattern as the in sample data. Another important point to note here is that the process governing the behavior of the tail varies as we change the confidence intervals of interest. The in-sample results show the precision of fit of the models. A strong fit suggests that the models have succeeded in describing the evolution of the left tail of the stocks. Again, a very high match of the percentage of hits does not necessarily translate to a better DQ test result.

Simulation studies are carried out to compare the performance of the CAViaR group of models with the SV linear predictor models (SVLIN, A-SVLIN) when the underlying data is from SV.

Jacquier *et al.* (1994) discuss parameter choices (for SV parameters, θ), driven by empirical findings of the daily returns and further verified by findings of several other researchers. Melino and Turnbull (1990) analyzed CD/$ exchange rate collected daily from 1975 to 1986 and obtained $\hat{\alpha}_1 = 0.91$ and coefficient of variation, CV= 0.14, Danielsson (1994) studied the daily S&P 500 (1980–1987) and recorded $\hat{\alpha}_1 = 0.96$ and CV= 0.34, Harvey *et al.* (1993) studied daily exchange rate data (1981–1985) and reported $\hat{\alpha}_1$ to be 0.958 to 0.995 with the CV ranging from 0.47 to 0.74. $E(\sigma_t^2) = 0.0009$ was used throughout the sampling experiment in Jacquier *et al.* (1994)'s study. If the simulated data are assumed to be weekly returns, this implies an approximately 20% annual standard deviation, typical of most stocks. The choice of α_1 is unanimously decided to be in the range of 0.9 to 0.995. The choice of the other parameters are driven by the fact that the coefficient of variation of the volatility process has been found to lie in the 0.1 to 1 range.

We conducted an experiment with recent data by applying Quasi maximum likelihood (details in Section 4.1 of Chapter 4) for parameter estimation and then obtained the coefficient of variation of the smoothed volatility estimates and came to a similar conclusion. 15 daily stock values are obtained from Yahoo! Finance from 7/86 to 6/08. The log

Table 2.13: *Summary Statistics of the log returns and coefficient of variation of the estimated volatilities*

	Mean	STD	Skewness	Kurtosis	CV
XOM	0.0583	1.4966	-0.8430	26.1481	0.1909
WMT	0.0569	1.8768	-0.0418	6.2208	0.3398
TXN	0.0528	2.8806	-0.0966	8.6148	0.3624
T	0.0427	1.6845	-0.1599	6.9291	0.3175
PFE	0.0431	1.8257	-0.3513	8.2139	0.2493
MRK	0.0451	1.7703	-1.3124	25.1690	0.1727
MER	0.0552	2.3970	-0.3583	12.8511	0.3176
KO	0.0504	1.6347	-0.7429	25.2609	0.3842
IBM	0.0308	1.8660	-0.5199	17.2651	0.3996
HPQ	0.0489	2.5128	-0.1666	10.0642	0.3530
GM	0.0061	2.1034	-0.1267	9.2315	0.2211
GE	0.0501	1.6616	-0.3020	11.0985	0.3527
F	0.0204	2.1378	-0.0355	7.7551	0.1423
DNA	0.0482	2.6540	1.0430	38.9355	0.8742
C	0.0438	2.2025	-0.5430	13.6440	0.3325

returns are calculated and any missing values are discarded. Summary statistics of the log returns and coefficient of variation of the smoothed volatility estimates are reported in Table 2.13. This validates using the parameter ranges used by Jacquier *et al.* (1994). For the simulation study, SV parameters $\theta = \{\alpha_0, \alpha_1, \sigma_v^2\}$ are chosen over a grid of values shown in Table 2.14, where $\mu_h = \alpha_0/(1-\alpha_1)$ and $\sigma_h^2 = \sigma_v^2/(1-\alpha_1^2)$ are the mean and variance of the

Table 2.14: *SV parameter values used in simulation*

	θ	α_0	α_1	σ_v	μ_h	σ_h^2
	θ_1	-0.736	0.9	0.363	-7.36	0.6935
CV=1	θ_2	-0.368	0.95	0.26	-7.36	0.6933
	θ_3	-0.146	0.98	0.166	-7.3	0.6959
	θ_4	-0.7061	0.9	0.135	-7.061	0.0959
CV=0.1	θ_5	-0.353	0.95	0.0964	-7.06	0.0953
	θ_6	-0.1412	0.98	0.0614	-7.06	0.0952

marginal distribution of the log volatility process. Because of the presence of high kurtosis seen in financial data, error distributions with heavy tails viz., t with 1 and 2 d.f. along with standard Normal are considered. Two percentages of VaR, 1% and 5% are considered

for analysis. Three thousand five hundred data points are generated from the SV model for each of the θ parameter cases. This is divided into two sets of data, 3000 data points used for estimation of the model comprising the in-sample data, and the remaining 500 is used for evaluation purposes, the out-of-sample data. The estimated parameters obtained based on the in-sample data is used to calculate the VaRs' and their Hit-in-sample and Hit-out-of-sample are recorded for each run. This is further used to obtain the DQ test results (both in-sample and out-of-sample). This is repeated for 1000 MC runs. The estimate statistics of the parameters based on the grand mean and variance-covariance matrix from all the runs are reported. The Hit-in-sample and the Hit-out-of-sample means along with their empirical C.I.s are measured. Finally, the proportion of the acceptance of the DQ tests based on their p-values at 1% level of significance is shown. The four CAViaR models with SVLIN are used for comparison. The models considered are all continuous and differentiable in β. Since the RQobjective function is a nonparametric objective criterion, it is used to check their performance for heavy tailed error distributions as well. Furthermore, VaR filtering from Watanabe's NFML algorithm provides a benchmark to compare the aforementioned models.

Tables 2.15 – 2.20 show the results of the best CAViaR model with the SVLIN model for the different error distributional assumptions and different parameter settings.

Tables 2.21 and 2.22 show the out of sample performance of the VaR forecasts with respect to the different methods. The instruments for the out of sample DQ test are a constant, the VaR forecast and the first four lagged hits. The formulae to compute $\hat{D}_T$ and $\hat{M}_T$ were implemented using k-nearest neighbor estimators, with k=40 for 1% VaR and k=60 for 5% VaR. The Nelder-Mead Simplex algorithm is used for the optimization.

Among the four models proposed by Engle and Manganelli (2004), the symmetric model as expected, perform better in modeling the dynamics when the underlying data are generated from an SV model with the given specifications. It might be pointed out that the Symmetric(SAV), Asymmetric and the iGARCH models can be looked upon as the quantile counterparts of the ARMACH models (Taylor, 1986), GARCH followed by standard deviation and with asymmetric errors (Nelson, 1991), and GARCH (Bollerslev, 1986) respectively with i.i.d. error distributions. The adaptive model has a unit coefficient on the lagged VaR and whenever the VaR is exceeded its value is increased. But that will not depend on the magnitude of exceedence. Hence, it is expected that it will slowly capture the dynamics with respect to the others, especially when the persistence is high.

Table 2.15: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 5% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_1 = \{-.736, 0.9, .363\}$ with coefficient of variation 1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.736, 0.9, 0.363)$					
	N(0,1)		t_2		t_1	
5% VaR	iGARCH	SVLIN	SAV	SVLIN	SAV	SVLIN
β_1	0.0002	-0.4674	0.0164	-0.4001	0.0914	-0.2557
Var	0.0000	0.0807	0.0004	0.3371	0.0129	0.5203
β_2	0.7348	0.8021	0.7417	0.8091	0.4653	0.8223
Var	0.0051	0.0040	0.0662	0.0194	0.4198	0.0536
β_3	0.4346	0.0871	0.1274	0.0698	0.0143	0.0510
Var	0.0139	0.0004	0.0068	0.0009	0.0051	0.0024
RQ	9.6232	9.7711	26.8854	26.7675	385.7361	385.5010
Hits in(%)	5.0179	5.0201	5.0010	5.0058	4.9993	5.0145
Emp. C.I.	(4.97, 5.07)	(4.97, 5.1)	(4.97, 5.07)	(4.97, 5.07)	(4.9, 5.1)	(4.9, 5.13)
DQ in acc%	99.7	89.2	99.2	99.2	98.6	99.8

Table 2.16: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 5% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_2 = (-0.368, 0.95, 0.26)$ with coefficient of variation 1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.368, 0.95, 0.26)$					
	N(0,1)		t_2		t_1	
5% VaR	iGARCH	SVLIN	SAV	SVLIN	SAV	SVLIN
β_1	0.0001	-0.1641	0.0083	-0.1242	0.0844	-0.0864
Var	0.0000	0.0215	0.0002	0.0711	0.0120	0.1934
β_2	0.8152	0.8691	0.8459	0.8758	0.5040	0.8725
Var	0.0021	0.0014	0.0307	0.0051	0.3967	0.0246
β_3	0.3700	0.0750	0.1135	0.0627	0.0146	0.0501
Var	0.0086	0.0003	0.0048	0.0005	0.0056	0.0020
RQ	9.3829	9.5555	26.7284	26.5709	373.7774	373.3412
Hits in(%)	5.0254	5.0170	4.9984	5.0071	4.9996	5.0167
Emp. C.I.	(4.97, 5.1)	(4.97, 5.1)	(4.97, 5.07)	(4.95, 5.1)	(4.9, 5.1)	(4.93, 5.1)
DQ in acc%	99.8	89	99.1	99.4	98.7	99.7

Table 2.17: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 5% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_3 = (-0.146, 0.98, 0.166)$ with coefficient of variation 1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.146, 0.98, 0.166)$					
	N(0,1)		t_2		t_1	
5% VaR	SAV	SVLIN	SAV	SVLIN	SAV	SVLIN
β_1	0.0011	-0.0123	0.0042	0.0093	0.0843	0.0093
Var	0.0000	0.0052	0.0001	0.0114	0.0128	0.1576
β_2	0.8928	0.9202	0.9104	0.9243	0.5179	0.9159
Var	0.0006	0.0005	0.0201	0.0014	0.3949	0.0178
β_3	0.1785	0.0562	0.0865	0.0490	0.0153	0.0422
Var	0.0014	0.0002	0.0027	0.0003	0.0060	0.0019
RQ	9.4030	9.5670	27.3246	27.1383	384.7304	384.0053
Hits in(%)	4.9991	5.011	4.9977	5.0056	5.0009	5.0175
Emp. C.I.	(4.93, 5.07)	(4.93, 5.1)	(4.93, 5.07)	(4.93, 5.1)	(4.9, 5.1)	(4.93, 5.13)
DQ in acc%	99.4	90	99.2	99.5	98.4	99.7

Table 2.18: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 1% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_4 = (-0.7061, 0.9, 0.135)$ with coefficient of variation 0.1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.7061, 0.9, 0.135)$					
	N(0,1)		t_2		t_1	
1% VaR	SAV	SVLIN	ASYMM	SVLIN	SAV	SVLIN
β_1	0.0258	-0.8895	0.1304	-0.5538	0.4894	0.2098
Var	0.0013	3.7504	0.0180	1.6472	0.5247	0.7990
β_2	0.5987	0.8063	0.3440	0.7708	0.4854	0.4699
Var	0.2511	0.1383	0.4060	0.1712	0.51	0.4024
β_3	0.1189	0.0159	0.2071	0.0214	0.1412	0.0377
Var	0.0103	0.0005	0.1959	0.0030	0.4901	0.0144
β_4			0.1871			
Var			0.1961			
RQ	2.4791	2.4830	12.5288	12.5872	94.346	92.0057
Hits in(%)	0.9987	0.9967	0.9980	1.0010	0.9978	1.0286
Emp. C.I.	(0.97, 1.03)	(0.9, 1.07)	(0.93, 1.07)	(0.93, 1.07)	(0.9, 1.07)	(0.93, 1.23)
DQ in acc%	98.8	96.3	99.8	99.3	99.1	99

Table 2.19: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 1% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_5 = (-0.353, 0.95, 0.0964)$ with coefficient of variation 0.1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.353, 0.95, 0.0964)$					
	N(0,1)		t_2		t_1	
1% VaR	SAV	SVLIN	ASYMM	SVLIN	SAV	SVLIN
β_1	0.0159	-0.5920	0.1264	-0.4853	0.4713	0.2178
Var	0.0008	2.5657	0.0187	1.5476	0.5197	0.7267
β_2	0.7383	0.8620	0.3676	0.7894	0.5084	0.4923
Var	0.1678	0.0930	0.4124	0.1607	0.4869	0.3878
β_3	0.1129	0.0167	0.2083	0.0226	0.1446	0.0385
Var	0.0083	0.0003	0.1903	0.0027	0.5198	0.0132
β_4			0.1911			
Var			0.1931			
RQ	2.4676	2.4762	12.5312	12.5874	92.2692	91.2795
Hits in(%)	0.9995	0.9989	0.9999	1.0002	0.9966	1.0347
Emp. C.I.	(0.97, 1.03)	(0.93, 1.07)	(0.93, 1.07)	(0.93, 1.07)	(0.9, 1.07)	(0.93, 1.18)
DQ in acc%	98.6	96.5	99.9	99.7	99.2	99

Table 2.20: *β estimate means, variances, Regression Quantile function value, for the best CAViaR specifications including the Symmetric Absolute value, Asymmetric slope, Indirect Garch, Adaptive with SVLIN, when $\epsilon_t \sim Normal(0,1)$, $\epsilon_t \sim t_2$ and $\epsilon_t \sim t_1$ for 1% Value at risk are shown. Data is generated from SV model with parameters chosen from the set $\theta_6 = (-0.1412, 0.98, 0.0614)$ with coefficient of variation 0.1. The sample size T=3000(in-sample), 500(out-of-sample) and MC loop of 1000 is used. Hits in sample proportions with the empirical C.I. at 5% significance level are reported. For evaluation, acceptance proportions of DQ in sample test results are shown with 1% level of confidence*

	$\theta = (-0.1412, 0.98, 0.0614)$					
	N(0,1)		t_2		t_1	
1% VaR	SAV	SVLIN	SAV	SVLIN	SAV	SVLIN
β_1	0.0101	-0.3138	0.1097	-0.4317	0.4714	0.2123
Var	0.0006	1.3250	0.0195	1.4561	0.51	0.7648
β_2	0.8238	0.9155	0.4506	0.8084	0.5080	0.5389
Var	0.1280	0.0493	0.4328	0.1464	0.485	0.3647
β_3	0.0967	0.0162	0.1699	0.0215	0.1455	0.0377
Var	0.0055	0.0002	0.1353	0.0027	0.5292	0.0138
RQ	2.4463	2.4579	12.5975	12.5789	90.0424	89.1305
Hits in(%)	0.9991	0.9982	0.9986	1.0003	0.9964	1.036
Emp. C.I.	(0.97, 1.03)	(0.93, 1.07)	(0.93, 1.07)	(0.93, 1.07)	(0.9, 1.07)	(0.93, 1.32)
DQ in acc%	99.1	96.4	99.8	99.5	99.4	99

Table 2.21: *Comparison of the out of sample test results of the 5% VaR estimates obtained for the best CAViaR model, SVLIN, and benchmark. Error distributions of N(0,1) and t_2 with different parameter settings with coefficient of variation 1 are used.*

5% VaR	N(0,1)			t_2		
θ_1	iGARCH	SVLIN	NFML	SAV	SVLIN	NFML
Hits out(%)	5.0004	5.05	5.2620	5.0760	5.0320	5.2500
Emp. C.I.	(3.2, 7.2)	(3, 7.6)	(3.6, 7.2)	(2.8, 8.2)	(2.8, 7.3)	(3.2, 7.2)
DQ out acc%	98	89	98	94.3	95.7	97
UC acc%	98	96.2	98	96.8	97.6	97.7
M acc%	99.7	98	100	99.3	99.3	100
CC acc%	98.9	96.3	100	97.4	98.4	98
θ_2	iGARCH	SVLIN	NFML	SAV	SVLIN	NFML
Hits out(%)	4.98	5.02	5.2660	5.004	5.01	5.2280
Emp. C.I.	(2.8, 7.2)	(2.6, 7.4)	(3.4, 7.4)	(2.8, 8)	(2.8, 7.4)	(3, 7.2)
DQ out acc%	98	88.2	98	95	95.1	98
UC acc%	98.3	96.3	97	97.4	97.5	99
M acc%	99.7	98.2	100	99.8	99.5	100
CC acc%	98.5	95.8	99	98.5	98.5	98
θ_3	SAV	SVLIN	NFML	SAV	SVLIN	NFML
Hits out(%)	5.0640	5.0236	5.3080	5.0260	5.0160	5.2520
Emp. C.I.	(2.8, 7.4)	(2.6, 8)	(3.6, 7.6)	(2.6, 7.6)	(2.8, 7.4)	(3, 7.4)
DQ out acc%	95	88	97.3	95.2	95.5	97
UC acc%	97.4	93.5	98.9	95.1	96.8	99.1
M acc%	99.7	98.4	99.9	99.6	99.9	100
CC acc%	98.1	94.2	99.3	96.5	97.7	99

Table 2.22: *Comparison of the out of sample test results of the 1% VaR estimates obtained for the best CAViaR model, SVLIN, and benchmark. Error distributions of N(0,1) and t_2 with different parameter settings with coefficient of variation 0.1 are used.*

1% VaR	N(0,1)			t_2		
θ_4	SAV	SVLIN	NFML	ASYMM	SVLIN	NFML
Hits out(%)	1.0660	1.48	0.9800	1.2220	1.7	0.9960
Emp. Quant	(0.2, 2.2)	(0.2, 6.7)	(0.4, 1.8)	(0.2, 3.7)	(0.2, 5.4)	(0.2, 2)
DQ out acc%	88.2	80	90	82.7	84.7	93
UC acc%	97.2	91	94.3	92.1	91.4	93.1
M acc%	99.899	99	100	95.7	98.8	100
CC acc%	98.78	93.4	100	93.69	94.6	100
θ_5	SAV	SVLIN	NFML	ASYMM	SVLIN	NFML
Hits out(%)	1.059	1.29	1.0080	2.1156	1.4704	0.9820
Emp. Quant	(0.2, 2.2)	(0.2, 3.7)	(0.4, 1.8)	(0.2, 3.5)	(0.2, 5.2)	(0.2, 2)
DQ out acc%	88.5	81.4	90	81.6	83.6	95
UC acc%	97.3	92	94.4	92.5	92.2	94
M acc%	99.6	99	100	96.23	99	100
CC acc%	98.9	94.6	100	94.4	95	100
θ_6	SAV	SVLIN	NFML	SAV	SVLIN	NFML
Hits out(%)	1.0460	1.0600	1	1.5660	1.3400	0.9940
Emp. Quant	(0.2, 2.2)	(0.2, 2.6)	(0.4, 1.8)	(0.2, 7.2)	(0.2, 3.6)	(0.2, 2)
DQ out acc%	88	82	90	85.8	84.7	93
UC acc%	96.1	93	94.5	93.5	92.2	94.5
M acc%	99.2	99.2	100	97.05	98.7	100
CC acc%	98.48	95.2	100	95.8	95	100

The SVLIN model produces comparable results as the other CAViaR models with respect to the DQ test performance in the normality case. However with heavier tailed distributions, such as t, with degrees of freedom 1 and 2, the performance of the DQ test significantly improves. The results provided above illustrate the various situations. When data are generated with normal distribution errors, among the CAViaR group only, ranked in order of preference are the symmetric absolute value, asymmetric, and igarch models. Symmetric model proves to be better competent than iGarch, because using the absolute returns is a stabler approach than using squares of returns. The asymmetric slope tends to overfit the data with improved insample results. However, this results in a reduced acceptance of the out of sample test results. The performance of the adaptive model is the worst in all cases.

In all the heavy tailed distribution cases, SVLIN is the best model for VaR 5% cases in terms of both insample fit and out of sample performance. When calculating the 5% VaR, the SVLIN model's performance has been found to be better than its 1% VaR cases. Also, with a higher coefficient of variation, CV=1, SVLIN is better with respect to the other competing models in heavy-tailed distribution situations. However, with lower CV, it has been found to be comparable but never better especially in the standard Normal error distribution case. One reason can be attributed to the form of the linear filter, where the logarithmic series of VaRs' and y_t's are considered. The y_t's are a zero mean process, hence taking the logarithm would make the filter tilt towards the negative axis. However, with heavier tailed distributions, the logarithmic scale makes it a stabler model with improved performance. However, it must be pointed out that the acceptance rates of the test results are very high for all the models.

The empirical C.I.s of the Hit out-of-sample means have the narrowest interval for the NFML scheme for all the cases. The SVLIN model fits very well into all the 1% VaR cases and for the 5% VaR, especially for the t distributions, as shown by the DQ in-sample test acceptance rates. In the case of 1% VaR for the t_1 distribution, the empirical C.I.s of the out-of-sample hits are large even for the best CAViaR model. For example, (0.2, 53.4) for θ_4, (0.2, 43.8) for θ_5, (0.2, 46.8) for θ_6 whereas those for the SVLIN model are (0.2, 2.4), (0.2, 2.6) and (0.2, 2.6) respectively. The nonlinear filtering scheme used as benchmark produces close to 100% acceptance rates for all test results with tighter empirical C.I.s in all cases considered. This is expected since that quantile estimates are obtained from the full conditional densities.

2.9 Summary

In this chapter, we propose two quantile specifications based on the symmetric and asymmetric SV framework. Empirical results show that they can be added to the CAViaR group of models. Their performance (based on fitting and test evaluation), when studied under different parameter settings and when applied to stock data, validate their use in VaR calculation. The model seems to be better adapted for heavy tailed distribution situations. It is important to note that the evolution of the tails vary with respect to the different levels of significance of the VaRs. Hence, some models might be better adapted than the rest for a particular situation. It is important to use an array of different quantile specifications and make an informed decision based on its evaluation by the different tests. Based on the current analysis, the SVLIN model seems to be very useful in the presence of leptokurtic data. These models provide simplified routes to obtain the quantile estimates directly, without evaluating the one step conditional densities at each step, however, our empirical study shows that they provide a good approximation. Quantiles obtained from the NF scheme also provide another attractive alternative for the monitoring process.

The next chapter is inherently built as an extension of the quantile regression (QR) modeling framework applied to SV models for estimation purposes. We showed in this chapter how the QR technique yields quantile estimates for the SV process and in the next chapter we focus on using this tool in the SV parameter estimation process. We develop a methodology motivated by the Efficient Method of Moments (EMM). We are interested in developing a robust strategy that can be used both for monitoring as well as estimation purposes, especially when dealing with leptokurtic data. We call our method Regression Quantile Method of Moments (RQMM).

Chapter 3

Regression Quantile MM (RQMM)

3.1 Introduction

Stochastic Volatility (SV) models are often used to model the conditional heteroscedasticity of financial time series data. In this chapter we are concerned with the estimation of SV model parameters. Our method is motivated by the Indirect Inference (II) (Gourieroux *et al.*, 1993), and Efficient Method of Moments (EMM) techniques (Bansal *et al.* (1993), Bansal *et al.* (1995), Gallant and Tauchen (1996)). These procedures are useful for models whose complexity make likelihood based inference difficult or impossible. In such cases, these tools employ a criterion based on an auxiliary model that approximates the true model behavior. The gradients of the likelihood ie., score equations of the auxiliary model form the criterion for the moment equations in EMM. The quantile regression framework is an increasingly important empirical tool allowing parsimonious model fitting to entire conditional distributions. With a quantile estimate it is possible to acquire knowledge about the changes in location, shape, spread of the distribution. Our proposed method, Regression Quantile Method of Moments (RQMM) uses the gradients of the Regression quantile objective function of the auxiliary model as the criterion. The use of the check function as the criterion makes the tool less sensitive to distributional assumptions and helps in developing a robust inference methodology. Since the method is based on the regression quantile structure of the auxiliary model, it yields robust estimates in the presence of outliers and in case of local misspecification of the true model. In financial time series, where data is leptokurtic with a latent time-varying stochastic volatility process, such robust methods are desirable and serves as a flexible tool for analysis.

The Generalized Method of Moments (GMM) was introduced in the econometrics

literature by Hansen (1982). The wide application of Method of Moments (MM) estimators in economics and finance is a consequence of the magnitude of research associated to the area. As opposed to the Maximum Likelihood (ML) approach, MM estimators are not sensitive to the statistical properties derived from the distributional assumptions. The maximum likelihood approach requires the correct specification of the underlying distribution of the data. In real world situations, the distributional assumption may not coincide with the truth, leading to biased inference.

In some cases, calculating the likelihood can in itself be computationally intensive. In such cases, the MM framework may provide a computationally viable solution to the problem. The GMM is a way of applying the minimum chi-squared criterion in the method of moments framework when the number of moment conditions exceed the number of parameters to be estimated. In some situations where direct use of GMM is infeasible, there exist some modifications of the GMM approach, exploiting the moment conditions. One of these is the simulated method of moments (SMM), and a closely related approach is the indirect inference (II).

Simulation method is a popular tool among researchers, which can be attributed to the tremendous growth in computational speed and viability. In situations where direct analytical moment calculations are infeasible, SMM or II methods are used. If the moments of the model of interest are obtained by simulation then the approach is SMM; in cases where moments are derived from an auxiliary model, the approach is known as indirect inference, II.

The SMM is an extension of the GMM. McFadden (1989) and Pakes and Pollard (1989) independently introduced the SMM in computing expected responses in the context of discrete response models. Examples of using SMM in the estimation context is seen in panel data methods of Pakes and Pollard (1989), McFadden (1989), and Duffie and Singleton (1993)'s use in the asset pricing model context. Other applications can also be found in the biostatistics literature. When the inference is indirectly based on the moment conditions of the auxiliary model that encompasses or closely approximates the true model, Gourieroux *et al.* (1993) coined the term indirect inference to describe the resulting method of moments approach. The EMM is a special case of II when the scores of the pseudo-loglikelihood function define the moment conditions.

Given the earlier groundwork on quantiles regression, we propose to apply the RQ objective criterion instead of the pseudo-loglikelihood criterion to derive the moment

conditions. Although this might cause loss of efficiency due to the loss of information in a correctly specified model, the gain in robustness poses itself as an interesting exercise. In the case of the normality assumption of the error distribution, the score equations can be seen as the normal equations obtained in OLS estimation. Similarly, using RQMM with median as the quantile of interest would give rise to the equations for the LAD estimation. It is a well known fact that LAD is a more robust technique than OLS. Also, RQMM becomes EMM when the auxiliary model distribution is an asymmetric Laplace density.

3.2 Indirect Inference

Efficient method of moments is a special case of indirect inference procedure, when the moment conditions are derived from the score equations of the auxiliary model. However, Gourieroux *et al.* (1993) describe indirect inference more generally, proposing that the criterion could be any well meaning (satisfying some technical conditions) function. One of the examples they propose is a M-estimator type criterion. The RQ objective criterion is an M-type estimator that can be used under the II umbrella. The following section elaborates the methodology of EMM and RQMM.

3.2.1 Efficient method of moments

Bansal *et al.* (1993), Bansal *et al.* (1995), Gallant and Tauchen (1996) introduced the EMM method, which describes a systematic approach to generating moment conditions for generalized method of moments (GMM) of the parameters of a structural model. This method can be implemented in all cases where the loglikelihood calculation of the structural model is tedious. The idea is to use scores of a density of an auxiliary model that closely approximates the true data generating process and taking the expectation under the structural model to define the GMM criterion. The auxiliary model is chosen such that the score equations have an analytical expression that can be computed easily. If the auxiliary model nests the structural model, then the estimator is as efficient as maximum likelihood. This estimator requires that expectations under a structural model can be computed by simulation, quadrature or directly by analytical expressions, although the likelihood calculations may not be straightforward.

This methodology is a special case of indirect inference proposed by Gourieroux *et al.* (1993). The score is the derivative of the log density of the auxiliary model with respect to the parameters of the auxiliary model. The moment conditions are dependent both on

the parameters of the auxiliary and the structural model. The parameter estimates of the auxiliary model are obtained from the quasi likelihood estimates. The parameter estimates of interest of the structural model are obtained by minimizing the GMM criterion. The moment calculation in the GMM criterion entails simulations performed under the true model such that the asymptotic bias in the auxiliary parameter estimate is corrected. The optimal weighting matrix depends only on the auxiliary model, that is easily computed. Therefore, the choice of the auxiliary model is crucial to the specification of the weighting matrix (especially, in the case of overidentifying conditions).

3.2.2 Application of II methods to SV model

The SV model is given by

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \log \sigma_t^2 &= \alpha_0 + \alpha_1 \log \sigma_{t-1}^2 + \sigma_v v_t \\ (\epsilon_t, v_t) &\sim N(0, I_2) \end{aligned}$$

The model parameters are denoted by $\theta = \{\alpha_0, \alpha_1, \sigma_v^2\}$. The following set of inequality restrictions $|\alpha_1| < 1$ and $\sigma_v > 0$ are imposed to ensure that y_t is stationary and ergodic and that the parameters are uniquely identified. The volatility process induces higher order moment dependence on y_t. α_1 measures the volatility persistence and is often found to be close to 1 in empirical data. The model generates a leptokurtic unconditional distribution especially if the choice of distribution for ϵ_t is heavy-tailed. Without loss of generality, the model assumes the absence of any time varying mean. At the outset, the asymmetric leverage effect has been sidestepped; these extensions can however be easily implemented.

The EMM approach is a culmination of the efficiency provided by the maximum likelihood approach coupled with the flexibility of the GMM method. The score equations of the likelihood function of the auxiliary model are used as the moment equations, to attain the efficiency of the maximum likelihood. In SV model case, due to the latent volatility process, the exact likelihood evaluation requires integration over the unobserved realizations of the state vector. In order to calculate the likelihood, integration needs to be carried out over T infinite dimensional state vectors, where T is the total size of the sample. The idea is to replicate the true data generating process by a careful choice of moment conditions, guided by data characteristics derived from the observations. Hence, EMM can gain near efficiency if the auxiliary model (or, proxy model) closely approximates the true

data generator. Working under the GMM setup, the efficiency argument is influenced by the crucial choice of the proxy model. It employs an auxiliary model that closely approximates the true model. Let y_1, y_2, y_3, ..., y_T denote the observed time series data while true model is characterized by a set of parameters θ_0. EMM is summarized in the following steps:

- Choose an auxiliary model with density $p_a(y_t|Y_{t-1}, \eta_0)$ such that it closely replicates the true data generating process $p(y_t|Y_{t-1}, \theta_0)$ where $Y_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$ and likelihood computation is feasible (for the auxiliary model parameter estimation). Dimension of η should at least be the same size as θ. In case of exact identification, the minimization criterion is not affected by the choice of the weighting matrix. Even if the auxiliary model is misspecified, White (1982) has deduced that under regularity conditions, the QMLE in the limit attains the quasi-true value of η. Maximizing the quasi-likelihood implies minimizing the Kullback-Leibler distance criterion. The score functions are as follows:

$$\frac{1}{T}\sum_{t=1}^{T} s_f(Y_t, \hat{\eta}_T)$$

  where $s_f(Y_t, \hat{\eta}_T) = (\partial/\partial\eta)\log p(y_t|Y_{t-1}, \hat{\eta}_T)$ denotes score function of the auxiliary model. One of the preliminary choices of the auxiliary model is the GARCH(1,1). Other alternatives are discussed in the next Subsection 3.2.3.

- Compute a consistent estimator of the asymptotic variance covariance matrix as the weighting matrix given by

$$\hat{V}_T = \frac{1}{T}\sum_{t=1}^{T} s_f(Y_t, \hat{\eta}_T) s_f(Y_t, \hat{\eta}_T)'$$

  This data dependent weighting matrix must be such that it converges to a symmetric nonnegative definite matrix in the limit. This is obtained directly in the QML step, avoiding the need for further computation of the weighting matrix during the second step estimation.

- Starting from a fixed parameter value θ and an arbitrarily large t, say N, the data series y_t is simulated from the true SV model, denoted by $\tilde{y}_n(\theta), n = 1 \ldots N$ and we find $\hat{\theta}$ that minimizes the following quantity

$$m_N(\theta, \hat{\eta}_T)' \hat{V}_T^{-1} m_N(\theta, \hat{\eta}_T)$$

where $m_N(\theta, \hat{\eta}_T) = \frac{1}{N}\sum_{n=1}^{N} s_f(\tilde{Y}_n(\theta), \hat{\eta}_T)$ is the sample moment at the fixed estimated pseudo parameter $\hat{\eta}_T$. Under the simulated data, the sample average of the score provides an estimate of the expectation of the moment conditions under the true model for N large. The law of large numbers guarantees that the sample moment converges to its population counterpart. This consequently can be viewed as a calibration tool to identify the true model parameters that match the observed and simulated moment conditions. The EMM estimator is obtained by minimizing the following GMM criterion:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}}[m_N(\theta, \hat{\eta}_T)' \hat{V}_T^{-1} m_N(\theta, \hat{\eta}_T)]$$

3.2.3 Choice of Auxiliary models in EMM

The near efficiency to efficiency argument is to a large extent dependent on the choice of moments based on the auxiliary model. Gallant and Long (1997) have shown that with a judicious selection of the proxy model, the quasi scores span the true score vector, which results in asymptotic efficiency. As the proxy data generator comes close to approximating the true data generator, EMM estimator will result in full asymptotic efficiency as MLE.

In order to explore the finite sample properties, Andersen *et al.* (1999) considered several fully parametric and semi-nonparametric score generators. Keeping all the generators as conditionally Gaussian, they use the various ARCH type specifications to model the conditional heteroscedasticity e.g., GARCH, EGARCH and use squared Hermite polynomial expansions (SNP densities, Gallant and Nychka (1987)) to account for nonnormality and time series structure in the innovation process. The SNP model is given as follows:

$$f(y_t|\Omega_{t-1}, \eta) = \frac{1}{\sigma_t} \frac{[P(z_t)]^2 \phi(z_t)}{\int_{-\infty}^{\infty} [P(u)]^2 \phi(u) du}$$

where $z_t = (y_t - \mu)/\sigma_t = \epsilon_t/\sigma_t$. $P(.)$ denotes the Hermite polynomial expansion terms; the square is taken to ensure the positivity of the expression. Otherwise, the normalization ensures that the probability density form integrates to 1.

3.2.4 Properties of EMM

Gallant and Tauchen (1996) show that under regularity conditions, the EMM estimator is consistent and asymptotically normal.

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{as} N(0, [D_\theta' V^{-1} D_\theta]^{-1})$$

where $D_\theta = \partial/\partial\theta' m(\theta, \eta)$. The asymptotic covariance matrix can be estimated consistently by

$$Cov(\hat{\theta}_T) = \frac{1}{T}\left[\frac{\partial m_N(\hat{\theta}_T, \hat{\eta}_T)'}{\partial\theta}\hat{V}_T^{-1}\frac{\partial m_N(\hat{\theta}_T, \hat{\eta}_T)}{\partial\theta}\right]$$

Model diagnostic tests such as specification tests and over-identifying tests can be implemented as in any GMM procedure.

3.3 RQMM method

The RQ (Regression Quantile) criterion can be set up in the form of estimating equations which can be incorporated into the EMM paradigm instead of the score functions. The RQ criterion is a replacement of the loglikelihood criterion in EMM above. Hence, the scores in the moment conditions are substituted with the gradients of the RQ criterion. This would lend some more flexibility to the methodology and therefore, acts as a robust procedure in the presence of outlier and misspecification. In this case, the choice of the proxy quantile regression framework for the RQ is crucial, in the sense that it should act as a good substitute for the true conditional quantile behavior. Therefore, analogous to EMM where the score equations of the auxiliary model play an important role to replicate the true conditional distribution dynamics, our method uses quantiles to nonparametrically replicate the true conditional distribution dynamics for SV. Since it has been widely established that the auxiliary model in EMM case is GARCH(1,1), or EGARCH(1,1), as a parsimonious first choice proxy model, their corresponding conditional regression quantile model can be derived, due to the deterministic nature of the volatility process in the ARCH group of models. The GARCH(1,1) yields the iGARCH(1,1) (indirect GARCH, should not be mistaken for integrated GARCH, IGARCH) as described in CAViaR models, Section 2.3. The other alternatives are discussed in the next Subsection 3.3.2.

3.3.1 Methodology Development

The RQMM methodology can be described as follows: when the optimum parameter estimates for SV model are reached, given that the auxiliary model approximates the true data model well, the estimated conditional quantiles for the proxy model will closely approximate the conditional quantiles for the SV, and hence this methodology can be analogously used to estimate the θ_0 of the SV model. Our method is summarized in the following steps:

- Choose an auxiliary model with density $p_a(y_t|Y_{t-1}, \eta_0)$ such that it closely replicates the true data generating process $p(y_t|Y_{t-1}, \theta_0)$ where $Y_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$. As a preliminary reference proxy model, let us consider GARCH(1,1). Since we intend to minimize the RQ objective criterion, we can write the gradients of the check function as a basis to write the moment equations. If we denote by q_t the conditional τ^{th} quantiles of the auxiliary model, then

$$\begin{aligned} \mathrm{Prob}(\mathrm{y_t} < \mathrm{q_t}|\Omega_{\mathrm{t}-1}) &= \tau \\ \Rightarrow E[I(y_t < q_t) - \tau|\Omega_{t-1}] &= 0 \end{aligned}$$

where $I(.)$ is the indicator function. In a typical VaR problem, τ ranges mostly between .95 to .99. However, we can construct a multiquantile function (τ across the full range[0,1]; taking the sum over the different quantile values), to get a better estimate of the SV parameters. Consider a sample of observations $y_1, y_2, \ldots, y_T$ generated from the following model:

$$y_t = q_t(\beta) + u_{\tau t} \qquad \mathrm{Quant}_\tau(\mathrm{u_{\tau t}}|\Omega_{\mathrm{t}-1}) = 0$$

Let the RQ criterion or "check function" ρ (preferred choice of notation is ρ, and RQ in this paper) be denoted by

$$\mathrm{RQ}_{\tau,\mathrm{T}}(\mathrm{u}) = 1/\mathrm{T}\sum_{\mathrm{t}=1}^{\mathrm{T}}(\tau - \mathrm{I}(\mathrm{u_t} < 0))\mathrm{u_t}$$

where τ^{th} quantile of u is 0. Henceforth, τ is supressed in the expression for $RQ_{\tau,T}$. In order to obtain an estimate of β, we solve:

$$\underset{\beta}{\mathrm{argmin}}\, 1/T\sum_{t=1}^{T}[\tau - I(y_t < q_t(\beta))][y_t - q_t(\beta)] \equiv \underset{\beta}{\mathrm{argmin}}\,\mathrm{RQ_T}$$

Defining the gradient of the RQ criterion as $\psi_T = \frac{\partial}{\partial\beta}\mathrm{RQ_T}$ which is treated as the moment equations in this case given by:

$$\psi_T = \frac{1}{T}\sum(I(y_t < q_t) - \tau)\frac{\partial q_t}{\partial\beta}$$

where $q_t = q_t(\beta)$ which is the quantile regression form. So in the RQMM case, score equations are replaced by ψ_T. The asymptotic equivalent of the criterion function

is known as the binding function (Gourieroux and Monfort, 1996). The quantile regression for a GARCH(1,1) model is given by an iGARCH(1,1) model:

$$q_t = \sqrt{\beta_0 + \beta_1 q_{t-1}^2 + \beta_2 y_{t-1}^2}$$

The quantile regression form has been derived for the auxiliary model from GARCH(1,1) with iid errors.

- Let S be the symmetric nonnegative weighting matrix in the minimization of the criterion. This needs to be a consistent estimator of the variance covariance matrix of $\psi(Y_t, \hat{\beta}_T)$. We define the weighting matrix by taking the outerproduct of the $\psi(Y_t, \hat{\beta}_T)$. Compute a consistent estimator of the asymptotic variance covariance matrix as the weighting matrix given by

$$\begin{aligned} \hat{S} &= \frac{1}{T}\sum_{t=1}^{T} \psi(Y_t, \hat{\beta}_T)\psi(Y_t, \hat{\beta}_T)' \\ &= \frac{1}{T}\sum_{t=1}^{T} (\tau - I(y_t < q_t(\hat{\beta}_T)))(\tau - I(y_t < q_t(\hat{\beta}_T)))\nabla q_t(\hat{\beta}_T)\nabla q_t(\hat{\beta}_T)' \end{aligned}$$

A HAC estimate using Bartlett's or Parzen's kernel is often used in such cases.

- Starting from a fixed parameter value θ and an arbitrarily large t, say N, the data series y_t is generated from the true SV model, denoted by $\tilde{y}_n(\theta), n = 1 \ldots N$, and we find $\hat{\theta}$ that minimizes the following quantity:

$$\hat{\theta}(S) = \underset{\theta}{\operatorname{argmin}}\, \psi_N(\theta, \hat{\beta}_T)' \hat{S}^{-1} \psi_N(\theta, \hat{\beta}_T)$$

where $\psi_N(\theta, \hat{\beta}_T) = \frac{1}{N}\sum_{s=1}^{N} \psi_T(\tilde{Y}_s(\theta), \hat{\beta}_T)$ is the sample moment at the fixed estimated pseudo parameter $\hat{\beta}_T$. The strong law of large numbers guarantees that the sample moment converges to its population counterpart. The above objective function will be minimized at the conditional quantiles closest to that of the SV model and yield the SV parameter estimates $\hat{\theta}$.

3.3.2 Choice of Auxiliary models in RQMM

Because of the unknown latent process, the derivation of the loglikelihood as well as that of any recursive conditional quantile calculation formula of the SV model is tedious. Given that the auxiliary models used in the case of EMM are well known, eg. GARCH,

EGARCH, a natural extension would be to derive their corresponding conditional quantile functions. The extensive study conducted in Andersen *et al.* (1999) shows the importance of choosing a parsimonious model, especially with experiments with small data sizes. In their study, fully parametric forms of GARCH and EGARCH models are used along with SNP models with the leading terms as GARCH or EGARCH to describe the conditional heteroscedasticity present in the series. The semi-nonparametric structures are used to account for the nonnormality and time series structure of the innovation process. In cases, especially when the sample size is small, using a nonparametric form of SNP may not even 'converge'. Hence, let us focus on the parsimonious quantile specifications. Because of the deterministic nature of the volatility process in the ARCH group of models, their corresponding conditional quantiles are derived as follows.

The quantile function for GARCH(1,1) has already been given above. The derivation of the quantile specification for EGARCH is shown as an example. As long as the volatility process is a recursive function based on the data and an estimate σ_0, and the error process is assumed iid, the extension is straightforward. Let us consider the EGARCH(1,1) model:

$$a_t = \sigma_t \epsilon_t, \qquad (1 - \alpha B) ln(\sigma_t^2) = (1 - \alpha_1)\alpha_0 + g(\epsilon_{t-1})$$

where

$$g(\epsilon_t) = \theta \epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)]$$

In case the error distribution is i.i.d. standard normal, the model for $\log \sigma_t^2$ is as follows:

$$(1 - \alpha B) \log(\sigma_t^2) = \begin{cases} \alpha^* + (\gamma + \theta)\epsilon_{t-1} & \text{if} \quad \epsilon_{\mathrm{t-1}} >= 0 \\ \alpha^* + (\gamma - \theta)(-\epsilon_{t-1}) & \text{if} \quad \epsilon_{\mathrm{t-1}} < 0 \end{cases}$$

where B is the back-shift operator, $\alpha^* = (1-\alpha)\alpha_0 - \sqrt{2/\pi}\gamma$. The conditional quantiles q_t's can be computed as follows:

$$\log q_t^2 = \alpha \log q_{t-1}^2 + g_1(y_{t-1}, q_{t-1})$$

$$g_1(y_{t-1}, q_{t-1}) = \begin{cases} \beta + \dfrac{\gamma + \theta}{c} \dfrac{y_{t-1}}{q_{t-1}} & \text{if} \quad \mathrm{y_{t-1}} >= 0 \\ \beta + \dfrac{\gamma - \theta}{c} \dfrac{-y_{t-1}}{q_{t-1}} & \text{if} \quad \mathrm{y_{t-1}} < 0 \end{cases}$$

Cornish-Fisher expansions are a natural extension to SNP Hermite polynomial densities. Cornish-Fisher (CF) expansions up to an order of a few terms for a standardized variate

can be easily derived. The final q_t estimate is obtained from the product of σ_t which is a GARCH or EGARCH and the standardized quantile obtained from the expansion. As derived in Lee and Lee (1992), and the algorithm used in Jaschke and Jiang (2001), a Cornish-Fisher expansion of a standardized variate also provides another option.

Again, a suitable choice for the truncation point in the CF expansion can be determined by an empirical exercise similar to Andersen *et al.* (1999)'s study for applications in the SV case and other situations. There is always a tradeof between a parsimonious and an overparametrized model. Overparameterization may lead to higher precision in some cases, but the pitfalls are lack of flexibility and in some cases, fitting to purely idiosyncratic noise in the data. From the applications perspective, this study is left for future analysis.

The other CAViaR specifications are more general and can be applied to non-iid data, and also in cases where the error densities and the volatilities are changing (Engle and Manganelli, 2004). The CAViaR specifications of the symmetric absolute value, proportional symmetric adaptive and the asymmetric slope model can be also be considered as the candidate proxy models. The notation $(x)^+ = max(x, 0); (x)^- = -min(x, 0)$

Symmetric absolute value: $q_t = \beta_0 + \beta_1 q_{t-1} + \beta_2 |y_{t-1}|$

Proportional symmetric adaptive: $q_t = q_{t-1} + \beta_1 (|y_{t-1}| - q_{t-1})^+ - \beta_2 (|y_{t-1}| - q_{t-1})^-$

Asymmetric slope: $q_t = \beta_0 + \beta_1 q_{t-1} + \beta_2 (y_{t-1})^+ - \beta_3 (y_{t-1})^-$

Asymmetric absolute value: $q_t = \beta_0 + \beta_1 q_{t-1} + \beta_2 |y_{t-1} - \beta_3|$

Third order symmetric: $q_t = \{\beta_0 + \beta_1 q_{t-1}^3 + \beta_2 |y_{t-1}|^3\}^{1/3}$

SVLIN: $\log q_t^2 = \beta_0 + \beta_1 \log q_{t-1}^2 + \beta_2 \log y_{t-1}^2$

Indirect ARGARCH $q_t = \beta_0 y_{t-1} + (1 - 2(I(\tau - .5)))(\beta_1 + \beta_2 (q_{t-1} - \beta_0 y_{t-1})^2 + \beta_3 (y_{t-1} - \beta_0 y_{t-1})^2)$

In order to get an unique solution, the dimension of the proxy model parameters should at least be the same dimension as that of the true model. In case of a dimension match, the estimate is independent of the choice of the weighting matrix.

A simulation study is conducted to gauge the relative performance of the aforementioned auxiliary models.

3.3.3 Misspecification Error due to Auxiliary Model

We denote the parameters defined in the auxiliary model as β and SV model parameters as θ. The true conditional τ^{th} quantile function (CQF), denoted by $Q_{t,\tau}(y_t|\Omega_{t-1}) = Q_{t,\tau}(y_t|X_t)$ depends on θ, where X denotes the conditional variables. For ease of notation τ is supressed henceforth. In case of GARCH(1,1), iGARCH(1,1) forms the regression quantile function; let the parameters in this form be denoted by $\beta = \{\beta_1, \beta_2, \beta_3\}$. β is a scalar multiple of η (auxiliary model parameters for EMM) in the GARCH, EGARCH models; however, it might not have an easily interpretable form in other situations. In Cornish Fisher expansions, they are related to the moments as derived from a Hermite density setup.

RQMM aims at finding an approximate proxy quantile regression that generates the proxy conditional quantiles as close as possible to the true conditional quantile function (CQF). Hence, a desirable property of minimizing the expected RQ objective function of the auxiliary model is to minimize the misspecification error caused by approximating the true conditional quantile function with a proxy. The criterion asymptotically is as follows:

$$\lim_{T\to\inf} \frac{1}{T}\sum_{t=1}^{T}(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta)) = E_\theta(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta))$$

The solution of the asymptotic problem is:

$$b(\theta) = \underset{\beta}{\operatorname{argmin}}\, E_\theta(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta))$$

where $b(\theta)$ is the binding function. In EMM, the maximization of the Quasi likelihood results in minimization of the Kullback Leibler information criterion that measures the proximity of the two conditional densities.

Minimizing the RQ for the conditional quantile of the auxiliary model aymptotically amounts to minimization of a weighted mean squared error loss function of the misspecification error (Angrist *et al.*, 2006). We state Angrist *et al.* (2006)'s theorem below and the proof is provided in the Appendix. Let us denote by $\Delta_\tau(X_t, \beta) = q_\tau(\beta) - Q_\tau(y_t|X_t)$, the misspecification error. The conditioning variables $X_t = \{Y_{t-1}, q_{t-1}, \ldots\}$ where $t = 1 \ldots T$, is supressed for ease of notation. The following theorem states that the proxy population conditional quantile minimizes the expected weighted mean square misspecification error given by $\Delta_\tau(X_t, \beta)$. Let ϵ_τ be the quantile specific residual, defined by the difference of the

response variable from the conditional quantile of interest: $\epsilon_\tau = Y - Q_\tau(Y|X)$ with density $f_{\epsilon_\tau}(e|X)$, at $\epsilon_\tau = e$.

Theorem 3.3.1. *Approximation Property (Angrist* et al.*, 2006): Suppose that*

(i) the conditional density $p(y_t|Y_{t-1}, \theta)$ exists a.s.,

(ii) $E[Y]$, $E[Q_\tau(y_t|X)]$, and $E(|X|)$ are finite, and

(iii) $b(\theta)$ uniquely solves the RQ objective function of the proxy model.

Then

$$\begin{aligned} b(\theta) &= \underset{\beta\in\Re^d}{\operatorname{argmin}} E_\theta[(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta))] \\ &= \underset{\beta\in\Re^d}{\operatorname{argmin}} E_\theta[(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta))] \\ &\quad -E_\theta[(\tau - I(y_t < Q_{t,\tau}(y_t|X_t)))(y_t - Q_{t,\tau}(y_t|X_t))] \\ &= \underset{\beta\in\Re^d}{\operatorname{argmin}} E[w_\tau(X,\beta)\Delta^2_\tau(X,\beta)] \end{aligned}$$

where

$$\begin{aligned} w_\tau(X,\beta) &= \int_0^1 (1-u) f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du \\ &= \int_0^1 (1-u) f_y(uq_\tau(\beta) + (1-u)Q_\tau(y_t|X)|X)du >= 0 \end{aligned}$$

The proof follows by showing that finding the argument minimum of the check function with respect to the auxiliary parameter is equivalent to minimizing the distance between the check functions of the true parameters from the proxy since the minimization is with respect to β. Thus the theorem shows that the population QR quasi-coefficient minimizes the weighted mean squared error, where the error is caused by approximating the true quantile with its proxy. The weights are given by the average density of the response variable over a line from the proxy conditional quantile $q_{t,\tau}(\beta)$, to the true conditional quantile, $Q_\tau(y_t|\Omega_{t-1})$. Premultiplication by the term $(1-u)$ in the integral results in more weight being applied at points on the line closer to the true CQF. The above theorem gives a justification for using the proposed RQ criterion.

The idea behind RQMM is to calibrate the estimate of β from the observed sample to the $\beta(\theta)$ obtained from the simulated sample, yielding the estimate for θ.

3.4 Asymptotic Properties of RQMM

The asymptotic properties of RQMM can be derived directly from the asymptotic results of the indirect inference estimator given in Gourieroux *et al.* (1993), Gallant and Tauchen (1996) and Gourieroux and Monfort (1996). The proofs and results follow in the RQMM case, where the criterion function is an M-type estimator.

After Koenker and Bassett (1978) introduced the regression quantiles models, a generalization of their linear model was proposed by Powell (1986) who considers the censored regression quantiles model. The large sample properties of the estimator are developed for error terms restricted to have a zero quantile, conditional on the regressors, with heteroscedasticity of unknown form, and demonstrates how to construct consistent estimators of the asymptotic covariance matrices. In the nonlinear case, Weiss (1991), White (1994) proves the consistency of the nonlinear regression quantile, both in i.i.d. and ergodic cases. Weiss (1991) shows consistency, asymptotic normality in a dynamic nonlinear model with non i.i.d. errors. Kim and White (2002) give consistency and asymptotically normality properties for $\hat{\beta}$ in misspecified models. We give here Kim and White (2002)'s results.

Let the number of parameters of β be denoted by q and that of θ by p($\leq$ q). Consider the auxiliary model as follows:

$$y_t = q(x_t, \beta_\tau) + \epsilon_{t,\tau} \qquad \text{Quant}_\tau(\epsilon_t|\Omega_{t-1}) = 0$$

Let ν_t denote the set of vectors determining the shape with associated shape parameter ϕ of the conditional distribution of the $\epsilon_{\tau,t}$. The error conditional density is given by $f_{\epsilon\tau}(u|X_t)$ as in the previous section. Let $u_t(\phi, \beta_\tau, s)$ denote the unconditional density of $s_t = (\epsilon_t, X_t, \nu_t)$. Let $\nabla \equiv \partial/\partial\beta$, $\nabla_i \equiv \partial/\partial\beta_i$, where β_i is the i^{th} element of β, $q_t(\beta) = q_t(x_t, \beta)$.

Theorem 3.4.1. *(Consistency of the auxiliary parameter: Kim and White (2002)) Under the regularity conditions CB1–CB6 (given in Appendix),* $\hat{\beta}_{\tau,T} - \beta_\tau = o_p(1)$ *where* $\hat{\beta}_{\tau,T}$ *is the unique solution of:*

$$\underset{\beta}{\text{argmin}}\, \tau \frac{1}{T} \sum_{t=1}^{T} [\tau - I(y_t < q_t(x_t, \beta))][y_t - q_t(x_t, \beta)]$$

The asymptotic normality derivation in the quantile regression literature (Powell (1984), Powell (1991), Weiss (1991)) builds on Huber (1967)'s theorem 3. Kim and White (2002) result is reported here.

Theorem 3.4.2. *(Asymptotic Normality of Auxiliary parameters $\hat{\beta}$: Kim and White (2002)) When the regularity conditions AN1–AN9 (given in Appendix) hold, and if the estimator is consistent, then:*

$$\sqrt{T}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N(0, D^{-1}A_0D^{-1})$$

where $A_0 = E[(\tau - I(y_t < q_t(x_t,\beta)))\nabla q_t(\beta)(\tau - I(y_t < q_t(x_t,\beta)))'\nabla' q_t(\beta)]$, *and* $D = E[f_t(0)\nabla q_t(\beta)\nabla' q_t(\beta)]$.

For the results that follow, let us introduce matrices J_0 and S_0 as follows:

$$J_0 = \operatorname*{plim}_T \frac{-\partial^2 RQ_T}{\partial\beta\partial\beta'}(y_T, X_T; b(\theta_0)),$$
$$S_0 = \lim V(\sqrt{T}\frac{\partial RQ_T}{\partial\beta}(y_T, X_T; b(\theta_0)))$$

Under standard regularity conditions, when interchange of differentiation and integration is allowed, J_0 is the same as D, and A_0 is S_0. D plays the role of the matrix of second derivatives and A_0, the outer product of the gradients. As defined earlier, $b(\theta_0)$ denotes the binding function.

$$b(\theta_0) = \operatorname*{argmin}_\beta E_\theta\{RQ(y_t|\Omega_{t-1};\beta)\} = \operatorname*{argmin}_\beta\{RQ_\infty(y_t|\Omega_{t-1};\beta)\}$$

The indirect inference estimator of θ is obtained by minimizing the following criterion:

$$\hat{\theta} = \operatorname*{argmin}_\theta \psi_N(\theta,\hat{\beta}_T)'\Sigma\psi_N(\theta,\hat{\beta}_T)$$

Following the results outlined in Gourieroux and Monfort (1996), we have the following theorem on consistency and asymptotic normality of $\hat{\theta}$.

Theorem 3.4.3. *Under the conditions T1–T5 given in Appendix, the RQMM estimator $\hat{\theta}$ is consistent and asymptotically normal.*

$$\lim_{T\to\infty}\hat{\theta} \to \theta_0 \quad a.s. \quad \text{in} \quad \text{Prob}$$
$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, W(\Sigma))$$

where $W(\Sigma) = [M_\theta\Sigma M_\theta]^{-1}M_\theta\Sigma S_0\Sigma M_\theta[M_\theta\Sigma M_\theta']^{-1}$, $M_\theta = \dfrac{\partial^2 RQ_\infty(\theta, b(\theta))}{\partial\theta'\partial\beta}|\theta = \theta_0, \beta = \beta_0$.

Corollary 3.4.4. *In the exactly identified case (p=q), M_θ is invertible, and* $W(\Sigma) = [\dfrac{\partial^2 RQ_\infty}{\partial\beta\partial\theta'}]^{-1}S_0[\dfrac{\partial^2 RQ_\infty}{\partial\beta\partial\theta'}]^{-1}$

The outline of the proof is given in the Appendix. The optimal choice for Σ is $\hat{S}^{-1}$, calculated from the outer product of the scores.

We investigate the properties of the RQMM estimator by a simulation study, discussed in the next section.

3.5 Simulation Study

Since RQMM is not sensitive to the choice of distributional assumptions and is proposed as a robust procedure, its inferential abilities are tested in the presence of outliers and local misspecification in the SV model with some heavy tailed nonnormal error distributions. The following candidate cases are considered for generating the distribution of ϵ_t with density $f_\epsilon(x)$:

- Contaminated Normal: $f_\epsilon(x) = I(x > \varepsilon)\phi(x) + I(x < \varepsilon)\frac{1}{K}\phi(\frac{x}{K})$, where $\varepsilon = 0.05$, $K = 10$ and $\phi(.)$ is the standard normal.
- t_ν: Two cases are considered where ν is 2, and 3,($\nu = 2$ producing the heavier tail)
- Data generated with fixed outliers: The y_t process is generated from SV, and is further mixed with a constant outlier $\zeta = 0.5$ as follows: $y_t' = (1 - h_t)y_t + h_t\zeta$, where $P(h_t = 1) = 0.02 = 1 - P(h_t = 0)$.
- Cauchy with location parameter 0 and scale as 0.5, 1, 1.5, 2, 2.5.

We are interested in assessing the robustness properties of this method with EMM in all these misspecification cases. The observed data size T is taken as 500. N=10000 is the simulation data size within each method. We report the bias and MSE of the parameter estimates obtained from 500 MC runs in Table 3.1, where CN stands for the Contaminated Normal case, and CO stands for the data with fixed outliers.

In most of the simulation studies conducted by Andersen *et al.* (1999), for different sample sizes, the GARCH model produced the lowest RMSEs. Hence, we use it as the reference auxiliary model for EMM in our study and compare its performance to RQMM method with indirect GARCH serving as the auxiliary model.

In case the structural model is correctly specified and known, the most efficient way of estimating the model is by maximum likelihood. However, as pointed out in Gallant and Long (1997), Gallant and Tauchen (1996), if the auxiliary model encompasses the true

model, EMM is as efficient, while if it is a good approximation, it is nearly as efficient asymptotically. Table 5 in Andersen *et al.* (1999) compares the bias, MSE of estimators using QML, GMM, EMM, MCMC, direct maximum likelihood calculations proposed by Fridman and Harris (1998), Watanabe (1999), Sandmann and Koopman (1998)'s MCL methods. EMM's performance lies close to the superior but highly cumbersome numerical methods of direct maximum likelihood procedures. Since there is a tradeoff between robustness and efficiency, a Monte Carlo study with 500 runs is conducted to compare RQMM and EMM under correct model specification, for T=500 and N=10000 and reported in Table 3.2.

As studied by Andersen *et al.* (1999), inference is sensitive to the choice of the auxiliary models in small sample sizes. Hence, a small sample study for sizes 50, 100 and 200 observations with 500 MC runs are carried out to evaluate the small sample biases and MSEs of EMM and RQMM estimates, reported in Table 3.3. The main purpose of the study is to however, gauge the performance of this method relative to EMM in terms of robustness. As is evident from extant literature on SV models, prior contributions by Andersen *et al.* (1999), Sandmann and Koopman (1998), Kim *et al.* (1998) and many other researchers in this area focus on the parameter setting used by Jacquier *et al.* (1994). Hence, we follow the gold standard setting for our evaluation. Parameter values of $\theta = \{\alpha_0, \alpha_1, \sigma_v\} = \{-0.7, 0.9, 0.363\}$ are used for the study. As suggested by Andersen and Sørensen (1996) and Andersen *et al.* (1999), in the second step, we use two series of length N which are identical, except that the first uses the innovation sequence $\{v_n, \epsilon_n\}$ and the other relies on $\{-v_n, -\epsilon_n\}$. This helps in mitigating the MC error in integration. Parsimonious models with GARCH leading term with the inclusion or exclusion of the Hermite polynomial approximation in SNP, has proven to be a strong candidate, especially for moderate sample sizes. Hence, they provide a good reference point for our simulation study. Finally, a comparison study between some choices of the auxiliary models discussed in Subsection 3.3.2 as competing models for RQMM is conducted for sample size T=500, N=10,000, and MC loops of 500, shown in Table 3.4. Andersen *et al.* (1999) has clearly pointed that in cases where a researcher is only concerned with the inference of the SV model, a direct likelihood based approach like Jacquier *et al.* (1994) MCMC, Kim *et al.* (1998) unified MCMC approach, Danielsson (1994) and Danielsson and Richard (1993)s' simulation based maximum likelihood (SML), Maximum likelihood Monte Carlo (MCL) of Sandmann and Koopman (1998), recursive numerical integration (ML) of Fridman and Harris (1998) and Watanabe (1999) may be the preferred approach; however, GMM and its variants

incorporate flexibility to the method. The intention of adding RQMM to this list is to add much more flexibility. The auxiliary models in RQMM is based on a general setup without a strong dependence on distributional assumptions, the data distribution calibration is based on a nonparametric and a possibly more robust construction. Hence, an assessment of RQMM performance in the aforementioned list of scenarios should serve to examine its potential application as a robust inference tool.

From Table 3.1, where cases of local misspecification with the error distributions following heavier tailed nonstandard specifications are considered, RQMM estimates show better accuracy and efficiency than EMM when evaluated in terms of the bias and MSE in almost all the cases. For illustration purposes, Figures 3.1 – 3.8 show the frequency density plots of the $\hat{\alpha}_1$ obtained from both methods. This further underlines the efficacy of RQMM in yielding robust statistics.

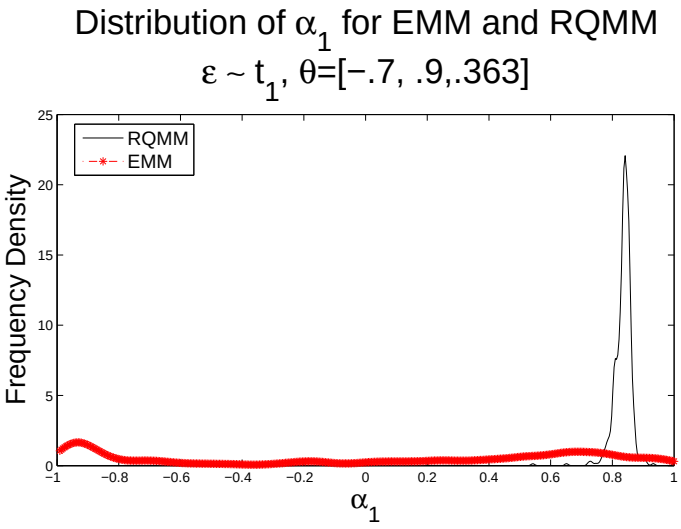

Figure 3.1: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim t_1$, *T=500, and 500 MC runs.*

Table 3.2 shows the comparison of RQMM and EMM estimates, where the underlying error distribution is standardized normal and in the absence of misspecification with moderate sample sizes of $T = 500$ and $T = 1000$. An interesting result is that the RQMM estimates for $\hat{\sigma}_v$ are found to be better than the EMM. RQMM produces better results than EMM which is not expected since the model is correctly specified in this case.

Table 3.1: *Comparison of Bias and MSEs for EMM and RQMM estimates for nonnormal distributions under local misspecification (where *(x)=(1.0e-003*x)). The sample size T=500, simulation size=10000(for integration of score) and MC run=500*

		RQMM			EMM		
		α_0	α_1	σ_v	α_0	α_1	σ_v
	Bias	-0.0050	-0.0184	0.0119	-0.0178	-0.0434	0.0134
t(2)	MSE	*0.4283	*0.4567	*0.2059	*6.1929	*3.0684	*0.9892
	s.t.d.	0.0201	0.0108	0.0081	0.0767	0.0344	0.0285
	Bias	-0.0058	-0.0106	0.0086	-0.0080	-0.0188	0.0094
t(3)	MSE	*0.2788	*0.2088	*0.1460	*1.7311	*0.5159	*0.5416
	s.t.d.	0.0157	0.0098	0.0085	0.0409	0.0128	0.0213
	Bias	-0.0042	-0.0056	0.0083	-0.0090	-0.0344	0.0153
CN	MSE	*0.1409	*0.0642	*0.1342	*1.9976	*1.7927	*0.4787
	s.t.d.	0.0111	0.0057	0.0081	0.0438	0.0247	0.0156
	Bias	-0.0068	-0.0101	0.0076	-0.0085	-0.0130	0.0053
DE	MSE	*0.5823	*0.2343	*0.3058	*0.8856	*0.2232	*0.7533
	s.t.d.	0.0232	0.0115	0.0158	0.0285	0.0074	0.0270
	Bias	-0.0094	-0.0010	0.0053	-0.0130	-0.0291	0.0130
CO	MSE	*0.2357	*0.0209	*0.0865	*2.3592	*1.0438	*0.4880
	s.t.d.	0.0122	0.0045	0.0076	0.0468	0.0141	0.0178
	Bias	-0.0000	-0.0203	0.0141	-0.0819	-0.5908	0.1313
Cauchy[0,.5]	MSE	*0.1942	*0.6224	* 0.2364	*16.7155	*815.3688	*39.5613
	s.t.d.	0.0140	0.0145	0.0062	0.1001	0.6836	0.1496
	Bias	-0.0130	-0.0677	0.0224	-0.1179	-0.8898	0.1970
Cauchy[0,1]	MSE	0.0005	0.0095	0.0006	0.0238	1.2958	0.0638
	s.t.d.	0.0189	0.0702	0.0118	0.0994	0.7107	0.1583
	Bias	-0.0249	-0.1073	0.0298	-0.1428	-1.1991	0.2657
Cauchy[0,1.5]	MSE	0.0010	0.0308	0.0020	0.0492	1.9455	0.0989
	s.t.d.	0.0195	0.1392	0.0332	0.1700	0.7132	0.1683
	Bias	-0.0311	-0.1363	0.0374	-0.1604	-1.3864	0.2996
Cauchy[0,2]	MSE	0.0013	0.0298	0.0021	0.0521	2.3090	0.1121
	s.t.d.	0.0172	0.1059	0.0270	0.1626	0.6226	0.1497
	Bias	-0.0448	-0.2025	0.0503	-0.1883	-1.4923	0.3311
Cauchy[0,2.5]	MSE	0.0031	0.0827	0.0041	0.0570	2.5781	0.1308
	s.t.d.	0.0331	0.2043	0.0401	0.1469	0.5931	0.1458

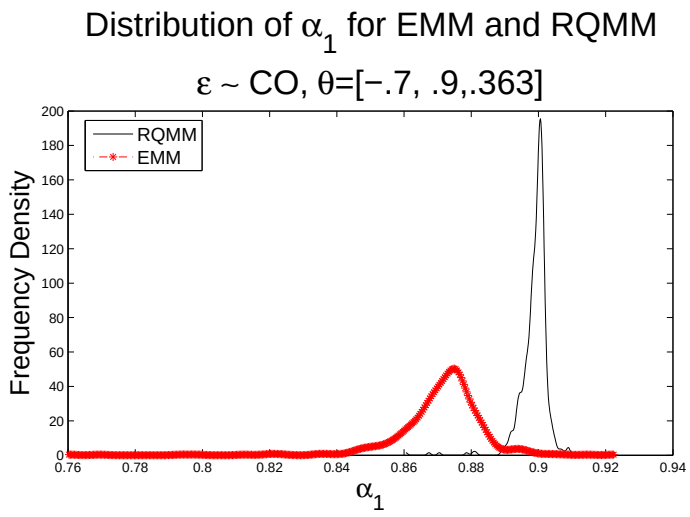

Figure 3.2: *Frequency Density plot of α_1 estimated by RQMM and EMM (dash-dot red line) when $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim CO$, T=500, and 500 MC runs.*

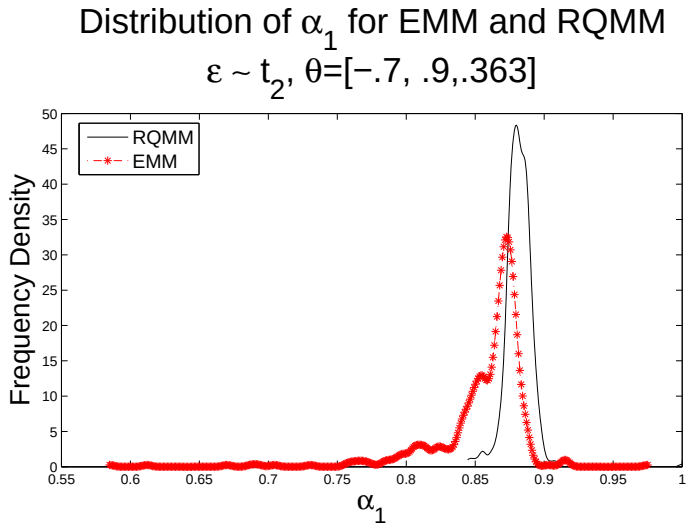

Figure 3.3: *Frequency Density plot of α_1 estimated by RQMM and EMM (dash-dot red line) when $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim t_2$, T=500, and 500 MC runs.*

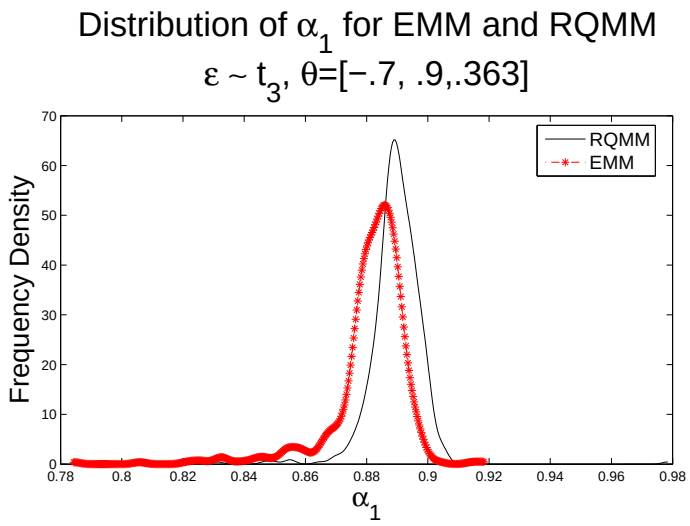

Figure 3.4: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim t_3$, *T=500, and 500 MC runs.*

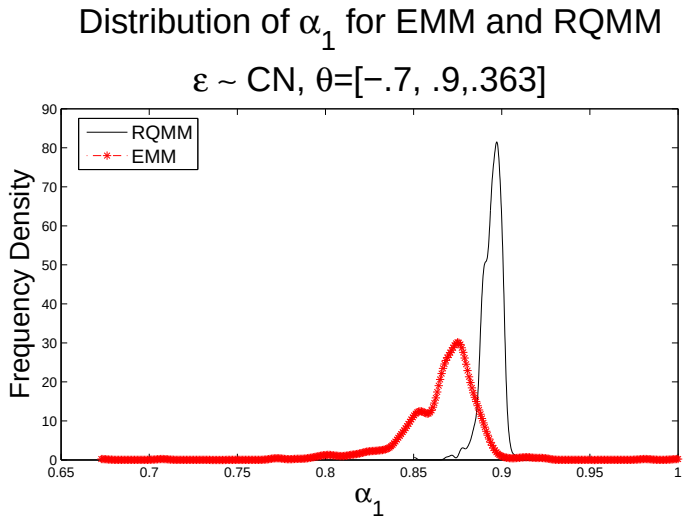

Figure 3.5: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim CN$, *T=500, and 500 MC runs.*

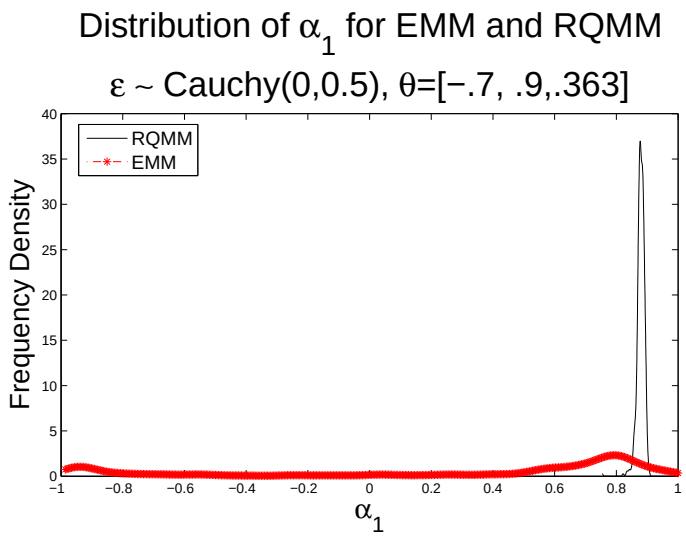

Figure 3.6: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim$ *Cauchy(0, 0.5), T=500, and 500 MC runs.*

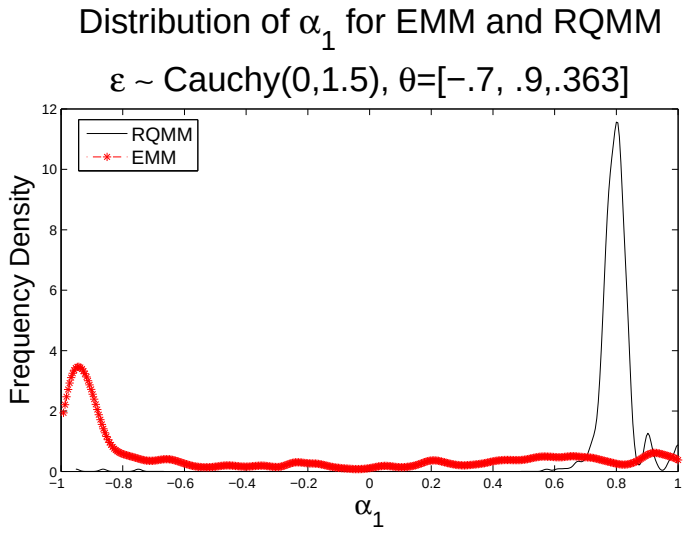

Figure 3.7: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim$ *Cauchy(0, 1.5), T=500, and 500 MC runs.*

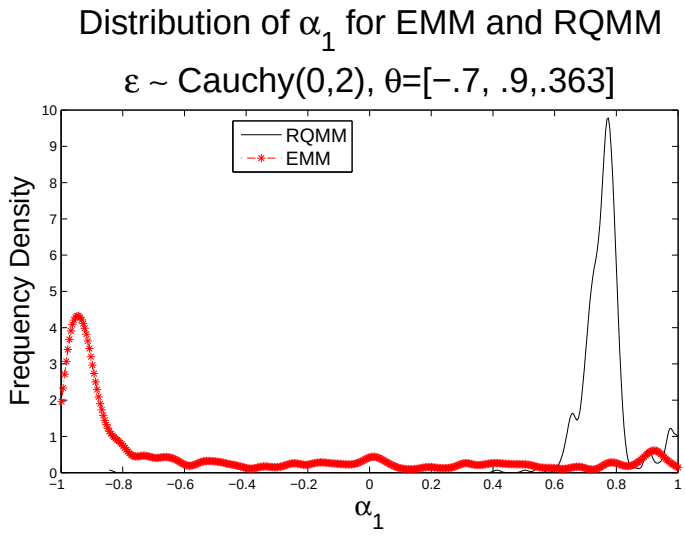

Figure 3.8: *Frequency Density plot of α_1 estimated by RQMM and EMM (dash-dot red line) when $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim$ Cauchy(0, 2), T=500, and 500 MC runs.*

Table 3.2: *Comparison of Bias, MSE and final estimates from RQMM and EMM, when the model has no misspecification, under the normality assumption with moderate sample sizes T=500, 1000, N=10000 and MC runs=500*

		RQMM			EMM		
		α_0	α_1	σ_v	α_0	α_1	σ_v
T=500	Mean	-0.7158	0.8989	0.3659	-0.7138	0.8987	0.3640
	Bias	-0.0158	**-0.0011**	0.0029	**-0.0138**	-0.0013	**0.0010**
	MSE	*0.5934	*0.0561	***0.2038**	***0.5847**	***0.0174**	*0.3878
	s.t.d.	**0.0186**	0.0074	**0.0140**	0.0199	**0.0040**	0.0197
T=1000	Mean	-0.7137	0.8984	0.3671	-0.7107	0.8987	0.3658
	Bias	-0.0137	-0.0016	0.0041	**-0.0107**	**-0.0013**	**0.0028**
	MSE	*0.4252	*0.0391	***0.1707**	***0.3502**	***0.0087**	*0.2342
	s.t.d.	0.0155	0.0061	**0.0124**	**0.0154**	**0.0027**	0.0150

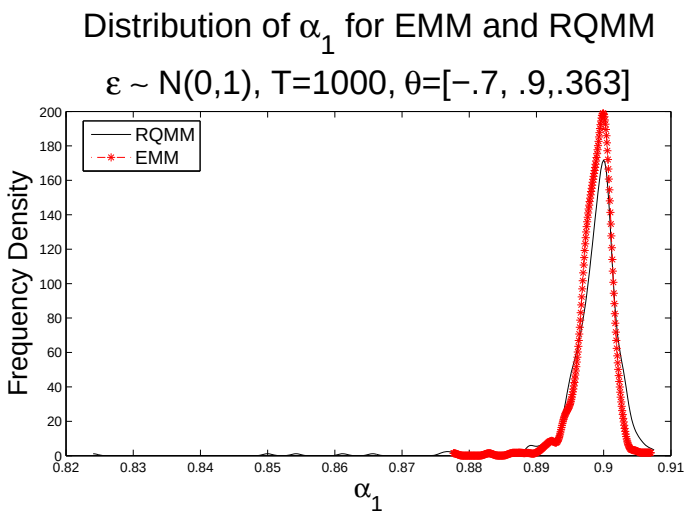

Figure 3.9: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim N(0,1)$, *T=1000, and 500 MC runs.*

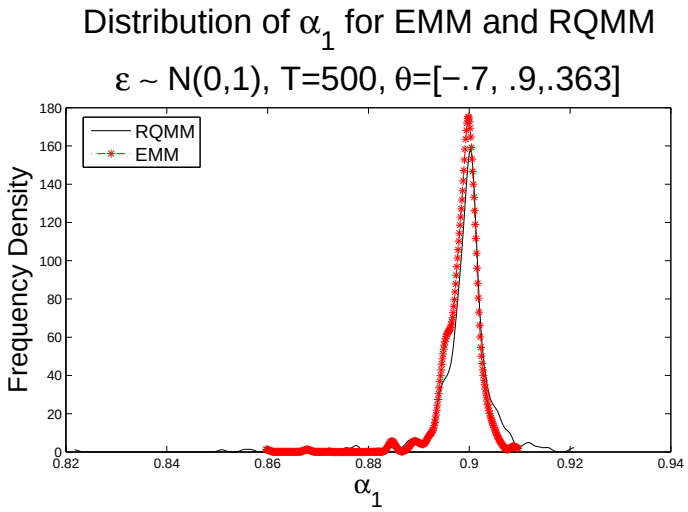

Figure 3.10: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim N(0,1)$, *T=500, and 500 MC runs.*

Table 3.3 shows the comparison of RQMM and EMM estimates, where the underlying error distribution is standardized normal and in the absence of misspecification with small sample sizes $T = 50$, $T = 100$, and $T = 200$. When there is no misspecification, for

Table 3.3: *Comparison of Bias, MSE and final estimates from RQMM and EMM, when the model has no misspecification, with T=50, 100, 200, N=10000 and MC runs=500*

		RQMM			EMM		
		α_0	α_1	σ_v	α_0	α_1	σ_v
T=50	Mean	-0.7156	0.9107	0.3613	-0.7227	0.8993	0.3613
	Bias	**-0.0156**	0.0107	**-0.0017**	-0.0227	**-0.0007**	**-0.0017**
	MSE	***0.6657**	*0.4609	***0.1472**	*1.0164	***0.3581**	*0.3216
	s.t.d.	**0.0206**	**0.0187**	**0.0120**	0.0224	0.0189	0.0179
T=100	Mean	-0.7200	0.9045	0.3625	-0.7234	0.8986	0.3584
	Bias	**-0.0200**	0.0045	**-0.0005**	-0.0234	**-0.0014**	-0.0046
	MSE	***0.8759**	*0.2129	***0.1322**	*1.9670	***0.1156**	*1.1911
	s.t.d.	**0.0218**	0.0139	**0.0115**	0.0377	**0.0107**	0.0342
T=200	Mean	-0.7181	0.9018	0.3628	-0.7216	0.8980	0.3599
	Bias	**-0.0181**	**0.0018**	**-0.0002**	-0.0216	-0.0020	-0.0031
	MSE	**0.0014**	0.0002	**0.0004**	0.0020	**0.0001**	0.0010
	s.t.d.	**0.0329**	0.0155	**0.0196**	0.0387	**0.0077**	0.0322

the small sample study, RQMM gives better estimates for α_0 and σ_v in terms of accuracy and efficiency evaluated by bias and MSE as demonstrated by the Figures 3.11 – 3.13 and Table 3.3. The better estimates are given in bold numbers.

3.6 Conclusion

In this chapter, we proposed a methodology based on II and QR tools to yield robust estimates in the presence of outliers and local mispecification of models. We illustrated the application of RQMM in nonstandard scenarios and compared their performance with respect to the established EMM procedure by conducting a simulation study. Based on the empirical results, we can conclude that RQMM is successful in controlling the bias and MSE in the presence of contamination.

Robustness often comes at the cost of efficiency. Therefore, in cases of correct specification of the model, RQMM is compared with EMM with respect to small and moderate sample studies. Our empirical findings suggest that the RQMM successfully yields robust statistics. The method gives an improved performance in small sample studies, especially

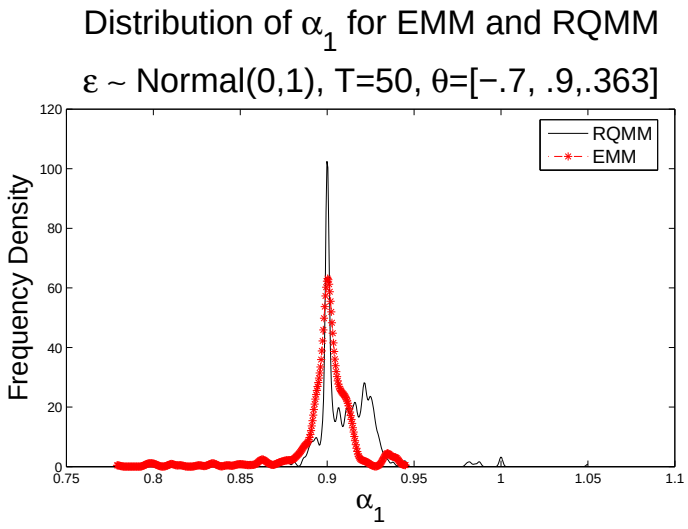

Figure 3.11: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim N(0,1)$, *T=50, and 500 MC runs.*

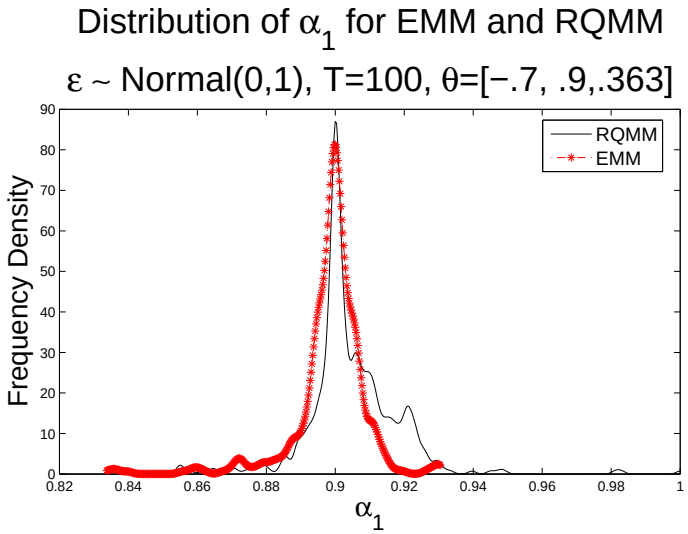

Figure 3.12: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$, $\epsilon_t \sim N(0,1)$, *T=100, and 500 MC runs.*

Table 3.4: *Comparison of Bias, MSE for competing Auxiliary models under RQMM with incorrect specification.* $*x$ *denotes* $1e^{-3} \times x$. *Errors are generated from t with 1 d.f.*

		RQMM		
		α_0	α_1	σ_v
SYMM	Mean	-0.7327	0.8307	0.3804
	Bias	-0.0327	-0.0693	0.0174
	MSE	0.0014	0.0234	0.0013
	s.d.	0.0196	0.1372	0.0324
ASYMM SL.	Mean	-0.7279	0.8303	0.3804
	Bias	-0.0279	-0.0697	0.0174
	MSE	0.0012	0.0211	0.0013
	s.d.	0.0198	0.1281	0.0311
SYMM-3	Mean	-0.7074	0.8854	0.3729
	Bias	-0.0074	-0.0146	0.0099
	MSE	*0.4827	*0.4834	*0.2151
	s.d.	0.0208	0.0165	0.0109
iGARCH	Mean	-0.7143	0.8346	0.3826
	Bias	-0.0143	-0.0654	0.0196
	MSE	0.0005	0.0067	0.0006
	s.d.	0.0165	0.0495	0.0132
SVLIN	Mean	-0.7465	0.6468	0.4084
	Bias	-0.0465	-0.2532	0.0454
	MSE	0.0096	0.3254	0.0114
	s.d.	0.0865	0.5137	0.0970
ASYMM ABS	Mean	-0.7508	0.7215	0.3871
	Bias	-0.0508	-0.1785	0.0241
	MSE	0.0144	0.1417	0.0019
	s.d.	0.1095	0.3331	0.0368
IAR-GARCH	Mean	-0.7120	0.8919	0.3694
	Bias	-0.0120	-0.0081	0.0064
	MSE	*0.3990	*0.9823	*0.1947
	s.d.	0.0161	0.0304	0.0125

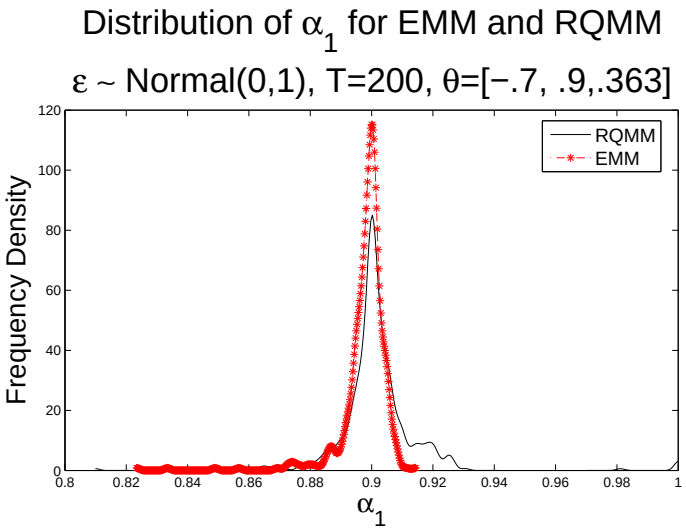

Figure 3.13: *Frequency Density plot of* α_1 *estimated by RQMM and EMM (dash-dot red line) when* $\theta = [-0.7, 0.9, 0.363]$*,* $\epsilon_t \sim N(0,1)$*, T=200, and 500 MC runs.*

for α_0, σ_v. For the memory parameter α_1, RQMM overestimates while EMM underestimates but is closer to the truth. MSE results are found to be comparable. In moderate sample sizes, when $T = 1000$, the estimates are comparable. The boxplots of the estimates show that the interquartile range contains the true parameter in most of the cases. Table 3.4 shows the performance of the competing auxiliary models under misspecification. All models produce estimates better than EMM. The simulation study clearly demonstrates the benefits of using this method for robust estimation purposes.

We considered data with high kurtosis in addition to the local misspecification of the model. Empirical findings verify the strength of this method. The data analysis in the previous chapter has already demonstrated the ability of the quantile specifications in fitting the tail behavior well. Hence, RQMM possesses the dual advantage of yielding robust parameter estimates of the SV model along with VaR calculation.

Chapter 4

Regression Quantile with Kalman filtering

In this chapter, we extend the Kalman filtering approach to obtain quantiles from a QR setup based on the SV model assumption. The state space model representation of the SV can be exploited by the Kalman filter to produce quantile estimates. The usage of the quantile regression (Koenker and Bassett, 1978) criterion can be looked upon as a further refinement over QML estimates. Most of the earlier work was related to model estimation with the primary focus on volatility forecasting. Our aim is to propose a method that serves the dual purpose of model estimation and VaR calculation directly.

In the next section, a brief outline of the Quasi Maximum likelihood (QML) is provided. This is followed by a description of the proposed method. For illustration purposes, simulation study results are presented in Section 4.3. Finally, the discussion is provided in Section 4.4.

4.1 Quasi Maximum Likelihood

A brief summary of the Quasi Maximum likelihood procedure developed by Nelson (1988), Harvey *et al.* (1993) and Ruiz (1994) is described. It uses the statespace form as mentioned earlier where we take the log of the squared terms in (4.1). Let y_t be the stochastic process of returns

$$y_t = \sigma_t \epsilon_t \qquad \epsilon_t \sim N(0,1) \tag{4.1}$$

where σ_t^2 is the conditional variance of y_t. In the simplest SV model framework, the log of volatility is expressed as an AR model:

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + v_t \qquad v_t \sim N(0, \sigma_v^2) \tag{4.2}$$

where ϵ_t and v_t are assumed to be independent of each other. The volatility is the latent variable, the parameters of the model are denoted by $\theta = (\alpha_0, \alpha_1, \sigma_v^2)$.

$$\ln(y_t^2) = \ln(\sigma_t^2) + \ln(\epsilon_t^2) \tag{4.3}$$

Let us define $Y_t = \ln(y_t^2)$, $x_t = \ln(\sigma_t^2)$ and $\eta_t = \ln(\epsilon_t^2)$. Hence (4.3) and (4.2) can be represented by

$$Y_t = x_t + \eta_t \tag{4.4}$$

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + v_t \tag{4.5}$$

The observation error η_t is a logarithmic chi squared random variable of degree 1 with mean and variance known to be approximately -1.27 and $\pi^2/2$ respectively. The standard Kalman filter is applied to (4.4) and (4.5). For a detailed description the reader is referred to Harvey (1989). The one step ahead prediction and updating formulae are as follows:
One step ahead prediction:

$$x_{t|t-1} = \alpha_0 + \alpha_1 x_{t-1|t-1} \tag{4.6}$$

$$P_{t|t-1} = \alpha_1^2 P_{t-1|t-1} + \sigma_v^2 \tag{4.7}$$

Updating:

$$x_{t|t} = x_{t|t-1} + \frac{P_{t|t-1}}{\delta_t}\eta_t \tag{4.8}$$

$$P_{t|t} = P_{t|t-1} - \frac{P_{t|t-1}^2}{\delta_t} \tag{4.9}$$

where $x_{t|t-1}$ and $P_{t|t-1}$ are respectively the optimal linear estimator and the variance of x_t given information till time $t-1$, and $x_{t|t}$ and $P_{t|t}$ are respectively the updated optimal linear estimator and the variance of x_t given information till time t. η_t and δ_t are respectively the prediction error and its variance given by

$$\eta_t = Y_t - x_{t|t-1} + 1.27 \tag{4.10}$$

$$\delta_t = P_{t|t-1} + \pi^2/2 \tag{4.11}$$

The optimal linear estimator means the estimator which minimizes the mean squared error (MSE) in the class of all linear estimators. Since the observational error η_t in our model is not normally distributed, the Kalman filter does not provide the minimum mean squared

error (MMSE) of the latent variable x_t although it still produces an estimator that minimizes the MSE among the class of linear estimators. We start the recursion by initializing the values $x_{1|0}$ and $P_{1|0}$ by using the knowledge that the unconditional mean and variance of x_t is known to be $\alpha_0/(1-\alpha_1)$ and $\sigma_v^2/(1-\alpha_1^2)$.

The QML is the likelihood obtained by assuming that the observational errors η_t are normally distributed. In that case, the likelihood is

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(\delta_t) - \frac{1}{2}\sum_{t=1}^{T}\frac{\eta_t^2}{\delta_t} \tag{4.12}$$

and can be maximized to obtain estimates of the SV parameters, θ. The QML estimate is often used as a preliminary estimate for other complex procedures for further refinement. We also use the QML estimates in our method as the initial method. The attraction of this method is its ease in implementation but as has been observed in Jacquier *et al.* (1994), they may not perform well in small sample sizes. The inefficacy can be attributed to the normality assumption.

The QR approach is a nonparametric approach that yields robust estimates. Since we are interested in obtaining VaR quantiles directly from our method, this tool is used to filter out quantiles and also yield SV model paramater estimates simultaneously.

4.2 Regression Quantile - Kalman Filter Method

The method of using the Regression Quantile (RQ) objective function (Koenker and Bassett, 1978) with a quantile model specification provides a robust approach for model estimation. To the best of our knowledge, there is no method that takes into account the VaR calculation directly in SV models, and also provide estimates of the SV model parameters simultaneously. Also, it may be noted here that the RQMM approach described in Chapter 3 is an intensive simulation-based procedure, whereas the method presented here is computationally simple. This proves to be an advantage, especially when time is a constraint. In this section, we propose a conjunction of the RQ objective function with the QML. Since the RQ criterion does not depend on any distributional assumptions, it is expected to provide robust estimates in the presence of outliers. The motivation of this method stems from the CAViaR model of Engle and Manganelli (2004). In the CAViaR group of models, the iGARCH quantile specification is derived under the GARCH assumption. The GARCH parameters can be directly estimated nonparametrically by estimating the quantile model using the RQ objective criterion. Although ML estimation produces

efficient statistics, preliminary simulation results as given in Engle and Manganelli (2004) show that those obtained using the RQ objective function (also referred to as the check function or tick function) yield good estimates.

In the SV model scenario, we were motivated to derive a quantile specification following a similar principle, since finding the exact likelihood itself is a challenging problem. Let the conditional quantile given the volatility process is known, be denoted by $q'_t(\tau)$, ie., $\tau = \int_{-\infty}^{q'_t(\tau)} P(y_t|\sigma_t)dy_t$. Given the true model is SV given by (4.1) and (4.2), the conditional recursive quantile structure is given by:

$$\ln {q'_t}^2 = \{\alpha_0 + \log(k^2)(1-\alpha_1)\} + v_t + \alpha_1 \log {q'_{t-1}}^2 \tag{4.13}$$

where $k = F^{-1}(\tau)$, where F is the CDF of ϵ_t.

The algorithm starts with obtaining estimates for v_t from the Kalman filter, $v_t = x_{t|t} - x_{t|t-1}$, computing the q'_t series recursively using (4.13), and finally refining the optimization with the help of the check function. But, in order to use the regression quantile function and estimate the parameters of the underlying model, the variables should undergo a scale transformation, such that it can be written in terms of all the parameters of the SV model. This technique can be seen as a plug-in method, where the estimates of $\hat{v}_t$ are plugged into (4.13), to obtain the quantile estimates. With the SV parameters embedded in (4.13), the regression quantile criterion can be used to estimate the parameters.

The equation (4.13) only contains the parameters $\{\alpha_0, \alpha_1\}$. Instead of using only the y_t returns process, we transform the variable to $z_t = \frac{\ln(y_t^2)}{\sigma_v}$. The transformed measurement equation become

$$z_t = x'_t + \eta'_t$$

where $x'_t = \dfrac{x_t}{\sigma_v}$, $\eta'_t = \dfrac{\eta_t}{\sigma_v}$, $\eta_t = \ln \epsilon_t^2$ and $\epsilon_t \sim N(0,1)$ as defined earlier. Further, the state space equation would be of the form:

$$x'_t = \alpha'_0 + \alpha_1 \frac{x_{t-1}}{\sigma_v} + v'_t$$

where $\alpha'_0 = \alpha_0/\sigma_v$, $v'_t = v_t/\sigma_v$. Hence, the new variate v'_t is standard normal. With the given framework, let us denote the ${x'}_{t|t-1}$ and ${P'}_{t|t-1}$ are respectively the optimal linear estimator and the variance of ${x'}_t$ given information till time $t-1$, and $x'_{t|t}$ and $P'_{t|t}$ are respectively the updated optimal linear estimator and the variance of x'_t given information

till time t. Therefore, $\hat{v}'_t = x'_{t|t} - x'_{t|t-1}$. With the q'_t as defined above, the recursive quantile model is therefore

$$\frac{\log q'^2_t}{\sigma_v} = \beta_1 + \beta_2 \frac{\log q'^2_{t-1}}{\sigma_v} + v'_t \tag{4.14}$$

where β_1, β_2 can be written in terms of the SV parameters as follows:

$$\beta_1 = (1 - \alpha_1)\frac{\ln k^2}{\sigma_v} + \frac{\alpha_0}{\sigma_v}$$

$$\beta_2 = \alpha_1$$

where k is as defined above by $k = F^{-1}(\tau)$. The q'_t obtained above are then used to minimize the RQ objective criterion given by:

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}[\tau - I(y_t < q'_t(\theta))][y_t - q'_t(\theta)]$$

The above minimization can be formulated to consider the percentiles or the deciles by summing over all the τ_is' (where eg., $\tau = (0.1, 0.2, \ldots, 1)$, in the case of deciles), in case the estimation problem is of interest only. The algorithm is as follows: Randomly generate 10,000 vectors from a Uniform(a,b) distribution to obtain initial parameter values. (a,b)'c range should contain the true parameter values. For the empirical study, Uniform random numbers from (-10,10), (0,1), (0,1) for $(\alpha_0, \alpha_1, \sigma_v)$ respectively, are used. We choose the first 10 $(\alpha_{0(0)}, \alpha_{1(0)}, \sigma_{v(0)})$ estimates with the largest log QML value. We use local optimization fminsearch (a Matlab inbuilt routine) to obtain 5 QML estimates. Use the standard Kalman filter to get the $v_t = x'_{t|t} - x'_{t|t-1}$. Plugging them in equation (4.14), the q'_t series is obtained. Finally, the check function is used as a final refinement and this optimization is repeated until convergence.

Once the parameter space is restricted to a local region using QML optimization step, the QR step with its robust loss criterion produces desired results.

In the out of sample case, where the q'_t are used for ex-ante VaR evaluation exercises, the estimates are obtained from the best linear predictor of past information of x'_ts' as opposed to the current information. In order to estimate k in (4.14) before the quantile recursion, smoothed volatility estimates can be obtained after the initial $\hat{\theta}$ values are obtained. These are used to obtain the empirical τ^{th} quantile based on the residuals, by standardizing the returns.

The method is computationally simple, and is implemented in heavy tailed error distribution cases eg., t_2, t_5 in the empirical study given below.

4.3 Empirical Study

4.3.1 Application to Stocks

In this section we will illustrate the application of RQ-KF to the stock data set. The RQ-KF method is tested with respect to the performance of the different out-of-sample tests. 5% and 1% VaR estimates are calculated using the optimal linear predictor values, based on the daily closing prices of DNA, GE, MER, F. The RQ objective criterion is minimized with respect to the percentiles during the estimation step. The Hits out of sample percentages were calculated by $\frac{1}{T}\sum_t^T I(y_t < -VaR_t) \times 100$, where T denotes the out-of-sample observations and $VaR_t = -q_t'$, following the convention used in reporting VaR values as a positive number. The results are presented in Table 4.1.

Table 4.1: *Comparison of the out of sample test results of the 5% and 1% VaR estimates obtained from RQ-KF for GE, F, DNA and MER*

	5% VaR			
	GE	F	DNA	MER
Hits out(%)	3.2025	5.1653	5.3719	5.7851
DQ out (p-val)	0.0204	0.4082	0.4224	0.0179
UC out (p-val)	0.0103	0.8135	0.5955	0.2624
M out (p-val)	0.0886	0.7890	0.5990	0.0155
CC out (p-val)	0.0055	0.9386	0.7588	0.0294
	1% VaR			
Hits out(%)	1.0331	1.2397	1.8595	2.1694
DQ out (p-val)	0.9285	0.7285	0.0508	**0.0004**
UC out (p-val)	0.9177	0.4536	0.0072	**0.0003**
M out (p-val)	0.6477	0.1350	0.4088	0.0798
CC out (p-val)	0.8961	0.2520	0.0398	**0.0014**

The results obtained from the out-of-sample statistics in Table 4.1 are similar to those obtained in Tables 2.1–2.12 in Chapter 2, in terms of acceptance of the model given by (4.14) with most of the competing CAViaR models at 1% significance level. The p-values that are rejected at 1% level of significance are reported in bold, in the table. This holds true especially in cases of F, DNA, GE. As pointed out in Chapter 2, the quantile evolution can vary at 1% and 5% levels of significance, which is observed in the case of DNA and GE. The number of exceptions (hits) is one of the deciding factors; however, as seen in the

previous results, the out-of-sample tests provide information about the characteristics such as independence and the past exceptions dependencies which should also be considered as indicators of a good model. The instruments used in the DQ test are a constant, and VaR forecast. An interesting aspect of the present results is that they show better conformity of the model in the 5% VaR case than 1%. This is in complete contrast to the findings in Chapter 2. One possible reason might be that we allot equal weights to all the percentiles in the estimation in this study. Minimization of the RQ objective criterion with respect to the τ^{th} quantile of interest (when $\tau = .01$) might result in better fitting. F at both the 1% and 5% cases, DNA at the 5% case, MER at the 5% case, GE at 1% case produce the best fits in terms of strong acceptance of the out-of-sample tests and also by the Hits-out-of-sample percentages. For example, in the GE 1% case, the Hits-out-of-sample percentage is 1.0331, which is significantly better than the exceptions produced by the CAViaR and SVLIN models. This is further supported by strong acceptance p-values of the tests. For visual illustration, the VaR estimates at 1% and 5% levels are plotted against the corresponding returns for four stocks considered in Figure 4.1.

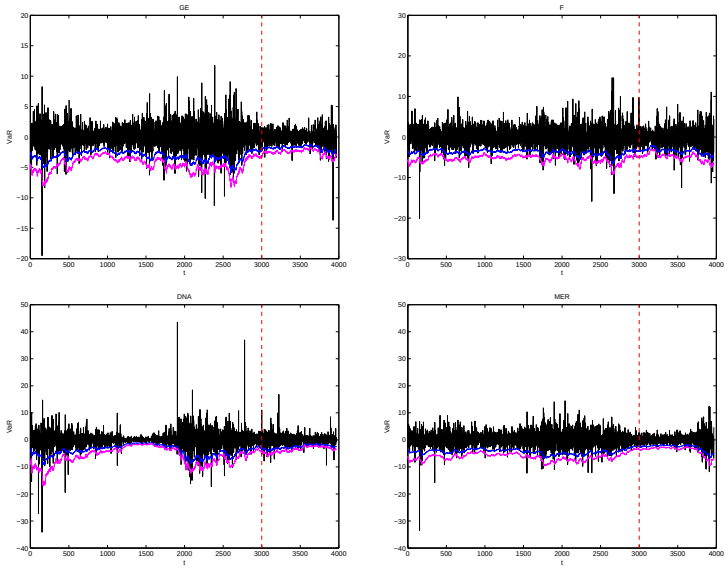

Figure 4.1: *5% and 1% VaR estimates for the short positions and their corresponding log returns. The red reference line serves to divide the out-of-sample from the in-sample data. The blue and magenta lines denote the* VaR_{95} *and* VaR_{99} *respectively.*

4.3.2 Simulation Study

A simulation study was carried out to compare the performance of the RQ-KF method and QML scheme in terms of estimation accuracy and efficiency. The RQ-KF method is tested on the same series of parameters used in Jacquier *et al.* (1994). Data are generated from the SV model with error distributions from standard Normal, t_2, and t_5 with total number of data points $T = 4500$. The first 3000 constitutes the in-sample data and the remaining serves as the out-of-sample data.

Using 1500 data for estimation by QML is a moderate sample size; larger datasets improves its performance. Since QML is known to show poor small sample performance, we are interested in the performance of the check function criterion over the initial results and its applicability as a tool for VaR assessment. Also, the time required for both the procedures (QML and RQ-KF) is relatively a lot less compared to any of the simulation based estimation procedures (RQMM, EMM, SMM, MCMC, etc.). A comparative study based on 500 MC loops is carried out between the QML estimates and RQ-KF estimates for illustration and the results are reported in Table 4.2.

The results obtained from Table 4.2 show that there is a slight improvement in the estimation of α_0 and α_1 with respect to the QML estimates in terms of bias, MSE. This is not true for the σ_v parameter though. With the use of box plots, as seen in Figure 4.3, the difference between the estimates obtained from both methods are not statistically significant. The MC interquantile ranges contain the parameter of interest. For illustration purposes, Figure 4.2, shows the plots of the α_1 estimates for the different distributional assumptions, and the three parameter settings considered in the simulation study.

In a notched box plot the notches represent a robust estimate of the uncertainty about the medians for box-to-box comparison. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level.

The estimation procedure is driven by the updated optimal linear filters $x'_{t|t}$. However, if we were to examine the quantile forecasts, the best linear predictor based on the previous information, i.e., $x'_{t|t-1}$ are used. The VaR estimates obtained are evaluated by the out of sample test performance, as illustrated in Table 4.3.

Table 4.2: *Comparison of Bias and MSEs' for RQ-KF and QML estimates for standard normal, t_5 and t_2 distributions. The sample size T=3000, and MC run=500*

		RQ-KF			QML		
θ_1		α_0	α_1	σ_v	α_0	α_1	σ_v
	Bias	**-0.0404**	-0.0073	0.0133	-0.0443	**-0.0060**	**0.0055**
t(2)	MSE	0.0673	0.0013	0.0114	**0.0590**	**0.0011**	**0.0057**
	s.t.d.	0.2565	0.0354	0.1060	**0.2391**	**0.0323**	**0.0754**
	Bias	**-0.0288**	-0.0053	0.0121	-0.0374	**-0.0051**	**0.0062**
t(5)	MSE	**0.0444**	**0.0008**	0.0094	0.0515	0.0009	**0.0050**
	s.t.d.	**0.2091**	**0.0287**	0.0965	0.2241	0.0303	**0.0703**
	Bias	**-0.0235**	**-0.0045**	0.0121	-0.0344	-0.0047	**0.0093**
N(0,1)	MSE	**0.0359**	**0.0007**	0.0085	0.0411	**0.0007**	**0.0041**
	s.t.d.	**0.1883**	**0.0256**	0.0914	0.1999	0.0269	**0.0637**
θ_2		α_0	α_1	σ_v	α_0	α_1	σ_v
	Bias	-0.0312	-0.0050	0.0106	**-0.0292**	**-0.0039**	**0.0052**
t(2)	MSE	**0.0210**	**0.0004**	0.0066	0.0244	**0.0004**	0.0034
	s.t.d.	**0.1418**	**0.0194**	0.0806	0.1537	0.0206	**0.0578**
	Bias	**-0.0183**	-0.0031	0.0139	-0.0206	**-0.0028**	**0.0051**
t(5)	MSE	**0.0115**	**0.0002**	0.0061	0.0145	0.0003	**0.0023**
	s.t.d.	**0.1057**	**0.0145**	0.0768	0.1189	0.0160	**0.0478**
	Bias	**-0.0260**	-0.0041	0.0152	-0.0263	**-0.0036**	**0.0091**
N(0,1)	MSE	**0.0112**	**0.0002**	0.0051	0.0121	0.0002	**0.0018**
	s.t.d.	**0.1026**	**0.0140**	0.0699	0.1068	0.0144	**0.0420**
θ_3		α_0	α_1	σ_v	α_0	α_1	σ_v
	Bias	-0.0151	-0.0023	0.0091	**-0.0075**	**-0.0010**	**0.0028**
t(2)	MSE	**0.0036**	**0.0001**	0.0032	0.0043	**0.0001**	**0.0010**
	s.t.d.	**0.0579**	**0.0079**	0.0557	0.0655	0.0088	**0.0323**
	Bias	-0.0096	-0.0015	0.0048	**-0.0000**	**-0.0000**	**0.0021**
t(5)	MSE	**0.0029**	**0.0001**	0.0022	0.0047	**0.0001**	**0.0010**
	s.t.d.	**0.0533**	**0.0073**	0.0465	0.0689	0.0093	**0.0317**
	Bias	-0.0163	-0.0024	0.0075	**-0.0087**	**-0.0012**	**0.0063**
N(0,1)	MSE	**0.0029**	**0.0001**	0.0020	0.0039	**0.0001**	**0.0009**
	s.t.d.	**0.0516**	**0.0071**	0.0441	0.0616	0.0083	**0.0286**

Table 4.3: *Comparison of the out-of-sample test results of the 5% VaR estimates obtained for the RQ-KF method. Error distributions of N(0,1), t_5 and t_2 and $T = 3000$ with different parameter settings are used.*

5% VaR			
θ_1	t_2	t_5	N(0,1)
Hits out(%)	5.4510	5.8421	6.0963
Emp. C.I.	(4.3, 6.7)	(4.5, 7.2)	(4.7, 7.6)
DQ out acc%	93.9	79.2	64.5
UC acc%	93.4	78.3	68.5
M acc%	98.3	96.9	94.8
W acc%	97.2	96.9	97.1
CC acc%	95.7	84.2	73.1
θ_2	t_2	t_5	N(0,1)
Hits out(%)	5.4506	5.7940	5.9991
Emp. C.I.	(4.2, 6.9)	(4.5, 7.2)	(4.5, 7.5)
DQ out acc%	92.3	79	66.9
UC acc%	93.4	79.6	70.1
M acc%	98.5	98.3	94.6
W acc%	97.2	95.6	96
CC acc%	94.2	86.1	74.3
θ_3	t_2	t_5	N(0,1)
Hits out(%)	5.4091	5.6615	5.8174
Emp. C.I.	(4.1, 6.7)	(4.2, 7.2)	(4.2, 7.4)
DQ out acc%	91.6	84.7	73.9
UC acc%	93.1	83.4	77.8
M acc%	97.6	98.5	95.6
W acc%	96.8	96.2	94.1
CC acc%	95.2	87.6	81.4

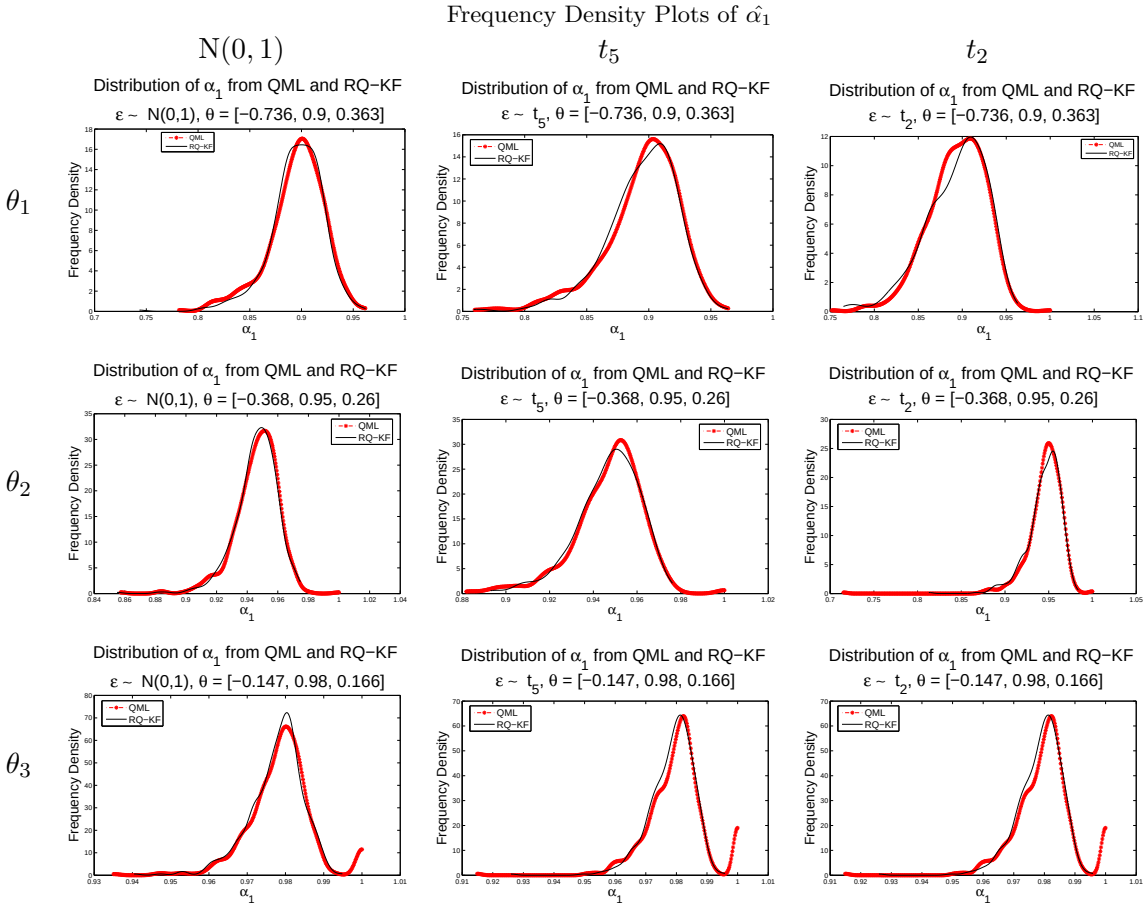

Figure 4.2: *Frequency Density plot of* $\hat{\alpha}_1$ *estimated by RQ-KF (black line) and QML (dash-dot red line) where* $\theta_1 = [-0.736, 0.9, 0.363]$, $\theta_2 = [-0.368, 0.95, 0.26]$, *and* $\theta_3 = [-0.147, 0.98, 0.166]$, *T=3000 from 500 MC samples.*

4.4 Conclusion

The main objective of the simulation study and data analysis presented in Section 4.3 was to evaluate the performance of RQ-KF method as a VaR computation tool as well as its ability to estimate the SV model. This is relevant in the context of SV models especially because its likelihood does not provide an analytical closed form solution and as a result, computation is not straightforward. The VaR calculation in practice depends on a two step method, where (a) volatility forecasts are obtained and (b) a tool is required to convert the volatility forecasts for VaR prediction. Our tool uses a quantile specification to directly estimate the model parameters as well as provide VaR estimates. Comparison with the

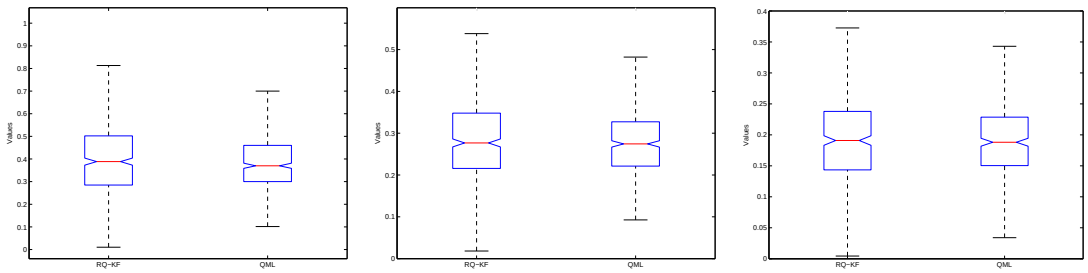

Figure 4.3: *Box plots of* $\hat{\sigma}_v$ *estimated by RQ-KF and QML. From the left panel,* $\theta_1 = [-0.736, 0.9, 0.363]$*,* $\theta_2 = [-0.368, 0.95, 0.26]$*, and* $\theta_3 = [-0.147, 0.98, 0.166]$ *are shown respectively.*

QML estimates show that the parameter estimates obtained are comparable based on the simulation results. Also, when RQ-KF is used to compute quantiles, application to stock data and simulation studies show that the tool performs well in the empirical evaluation studies as well. Hence, this methodology provides a simple and quick technique for model estimation and filtration of quantiles which can be further used for testing and evaluation.

As the α_1 parameter $\rightarrow 1$, and the $\sigma_v \rightarrow 0$, the performance of the method is expected to improve. This is confirmed by the results obtained in Table 4.3. With an increase in the sample size, the accuracy of the results increases, proven by the rate of acceptance of the out-of-sample tests. Our empirical findings suggest that summing the RQ objective function over the percentiles provides a better criterion than using lesser quantiles, such as deciles, in terms of fit. RQ-KF's performance in heavy tailed distribution cases is in particular, highly encouraging. Hence, this tool has its usefulness in the presence of high kurtosis prevalent in financial data. With a large data set, QML provides better initial estimates which help produce better quantile estimates.

In order to use the Kalman filter, a restriction of this method is that it can be applied to models that can be transformed to a state space form. In this respect, RQMM is a far more generalized tool and its advantage lies in its applicability in all situations. The simplicity and ease of application makes the RQ-KF procedure attractive, especially when time is a constraint. The check function adds a certain amount of flexibility especially when working with nonstandard cases. This method can be extended to the asymmetric SV models.

Chapter 5

Non-linear Filtering based on a Hermite polynomial approach

The difficulty in the estimation of the SV group of models arises from the addition of an innovation process in the volatility equation. Calculation of the exact likelihood requires integrating with respect to an infinite dimensional state vector and thus cannot be solved analytically. The exact likelihood is given by

$$\begin{aligned} L &= p(\theta|\{y_t\}_{t=1}^T) \\ &= \int\int\dots\int p(\{y_t\}_{t=1}^T|\{\sigma_t^2\}_{t=1}^T)p(\{\sigma_t^2\}_{t=1}^T|\theta)d\sigma_1^2 d\sigma_2^2\dots d\sigma_T^2 \end{aligned}$$

where $p(.|.)$ is the conditional density and T is the total number of observations.

In order to solve the infinite dimensional problem, while carrying out the integration in the nonlinear filtering scheme, Watanabe (1999) uses a judicious choice of nodes based on Tanizaki's (1993) technique. This is used for calculating the different probability estimates in the recursion steps of the NFML scheme. These can be used in computing the likelihood during the estimation process, or to find the VaR estimates in the monitoring process, as the case may be.

We interpret the NF scheme differently in terms of the moments as a basis of the recursion steps instead of the node points. This becomes feasible since a Gram Charlier (GC) representation of $p(x_t|\mathbf{y_{t-1}})$ exists. Our study shows that it can be defined effectively by its first few moments. Hence, the GC density based on Hermite polynomials is used as an effective replacement to the unknown conditional probability densities in NFML. The choice of density is facilitated by the assumption that the unconditional distribution of x_t follows

a Normal distribution. Therefore, the aim of our study was to seek a parsimonious fitting based on the Hermite polynomial coefficients. This exercise ultimately results in seeking an alternative finite-dimensional approach that uses a parametric form for the unknown densities.

Our empirical findings show that the first five hermite polynomials are enough to obtain the desired results in the SVAR(1) model. The five coefficients acts as a replacement for the 50 nodes used in Watanabe's method, resulting in reduced computational burden.

A brief outline of the nonlinear filtering method is given. The application of Hermite polynomials in this context is discussed next, followed by empirical results. A scope for future study using the Cornish Fisher expansion is discussed.

5.1 The Nonlinear Filter

The non linear filtering method of Watanabe (1999), Fridman and Harris (1998) uses the Bayesian rule in recursively calculating the conditional densities of the latent volatility given previous information on the return series and updating at every step, similar to the Kalman filter. The conditional return probability is obtained as a consequence of this filtering scheme. Each step requires calculation of an integral. Watanabe (1999) uses the trapezoidal rule and Fridman and Harris (1998) uses the Gauss-Legendre numerical integration rule. The SV parameters estimates are obtained by maximum likelihood. For details on the method, see Chapter 2, Section 2.7. An outline of the algorithm is as follows: Following previous notations,

- Start the calculation with $p(x_t|\mathbf{y_{t-1}})$, where $p(x_1|y_0) = p(x_1)$ is known.
- Calculate $p(y_t|\mathbf{y_{t-1}})$, using $p(y_t|x_t)$(known), and $p(x_t|\mathbf{y_{t-1}})$ from the previous step.
- Calculate $p(x_{t+1}|\mathbf{y_t})$ based on $p(x_{t+1}|x_t)$ (known), $p(y_t|x_t)$ (known), and $p(x_t|\mathbf{y_{t-1}})$.

These three steps outline the recursion steps.

Hermite polynomials

Hermite polynomials of degree n denoted by $H_n(x)$ is given by

$$H_n(x)\phi(x) = (-1)^n\phi_n(x)$$

where $\phi(x)$ is the standard normal density. $\phi_n(x) = \frac{d^n}{dx^n}\phi(x)$ is the n^{th} derivative of $\phi(x)$. The GC form of a density is given as follows:

$$p(y) = \sum\nolimits_{k=0}^{\infty} a_k H_k(y) w(y)$$

where $w(y)$ is the base density, and a_k can be written in terms of the moments as follows:

$$a_k = \frac{1}{k!}(\mu'_k - \frac{(k)_2}{1!2}\mu'_{k-2} + \frac{(k)_4}{2!2^2}\mu'_{k-4} - \frac{(k)_6}{3!2^3}\mu'_{k-6} + \ldots)$$

The orthogonality result of the Hermite polynomials, is,

$$\begin{aligned} \int_{-\infty}^{\infty} H_m(x)H_n(x)\phi(x)dx &= m! \quad \text{if} \quad \text{m} = \text{n} \\ &= 0 \quad \text{if} \quad \text{m} \neq \text{n} \end{aligned}$$

The implementation of the hermite polynomials in the recursion is discussed in the next section.

5.2 Approximation scheme

We propose an approximation scheme to the nonlinear filtering (NF) approach. Instead of finding an estimate of the pdf $p(x_t|\mathbf{y_{t-1}})$ at every step, we update the coefficients of the GC expansion of the pdf at each step. If a small number of terms is sufficient for the GC representation, the computation time is reduced.

In the NF scheme, the only unknown probability density function that appears in the integrand is $p(x_t|\mathbf{y_{t-1}})$. We approximate this density in terms of the GC density approximation. The hermite polynomials are an appropriate choice because of their domain on the $\Re$ line. This technique should in principle be adapted to all SV model extensions. We include some results on HSV models. If an appropriate approximation exists and the number of polynomial terms needed is decided apriori, then the entire algorithm reduces to finding the coefficients of the GC expansion. Hence, with every time step, instead of integrating over all the latent state variables, only the coefficients (number decided apriori) needs to be updated. This leads to dimension reduction, given that the number of coefficients are appreciably less than the number of latent state variables used.

Since we are interested in the recursive calculation of the coefficients, it eventually reduces to a linear regression fit problem. In case the base distribution is close enough to the true density, a GC expansion with a few terms proves to be quite efficient.

5.3 Details of the Approximation scheme

The step by step algorithm is:

- With the representation

$$p(x_t|\mathbf{y_{t-1}}) = \sum_{k=1}^{K} c_k^{(t)} H_k\left(\frac{x_t}{\sigma_h}\right)\frac{1}{\sigma_h}\phi\left(\frac{x_t-\mu_h}{\sigma_h}\right),$$

$K = 4$ is decided apriori and is fixed. Start the calculation taking $\vec{c} = [1,0,0,0,0]$ since $p(x_1|y_0) = p(x_1)$ is known. Given that we adopt a rule for choosing x_ts by the Kalman filtering technique described earlier, x_ts are known at the start of the recursion. Solving for the coefficients reduces to a linear regression problem. Create the matrix Δ with components $\Delta_{i,j} = \frac{1}{\sigma_h}\phi\left(\frac{x_t^{(j)}-\mu_h}{\sigma_h}\right) H_i\left(\frac{x_t^{(j)}}{\sigma_h}\right)$. Denoting $\Gamma = (\Delta'\Delta)^{-}\Delta'$. Therefore

$$c_k^{(t)} = \sum_{j=1}^{n} \Gamma_{i,j}^{(t)} p(x_t^{(j)}|\mathbf{y_{t-1}})$$

.

- Calculate $p(y_t|\mathbf{y_{t-1}})$, using $p(y_t|x_t)$(known), and $p(x_t|\mathbf{y_{t-1}})$ from the previous step.

- Calculate $p(x_{t+1}|\mathbf{y_t})$ based on $p(x_{t+1}|x_t)$ (known), $p(y_t|x_t)$ (known), and $p(x_t|\mathbf{y_{t-1}})$. Therefore, updating the coefficients:

$$c_i^{(t+1)} = \sum_{j=1}^{n} \Gamma_{i,j}^{(t+1)} p(x_{t+1}^{(j)}|\mathbf{y_t})$$

$$= \sum_{j=1}^{n} \Gamma_{i,j}^{(t+1)} \sum_{k=0}^{K} c_k^{(t)} \frac{\int p(x_{t+1}^{(j)}|x_t) p(y_t|x_t) \frac{1}{\sigma_h}\phi\left(\frac{x_t-\mu_h}{\sigma_h}\right) H_k(\frac{x_t}{\sigma_h}) dx_t}{p(y_t|\mathbf{y_{t-1}})}$$

Due to the discrete nature of the calculations, standardize the density estimates at every step.

Cornish Fisher (CF) expansions are often used to approximate the quantiles in finance. This would require finding the coefficients of the GC form of $p(y_t|\mathbf{y_{t-1}})$. Using a similar technique as described above, finding coefficients $d_l^{(t)}$ would entail forming a grid of points covering the range of observations y_t. This, however, does not result in a reduced

computation. If a risk manager already uses a Cornish Fisher algorithm for VaR calculations, the coefficients obtained from our algorithm can be used. We show that the first four coefficients are sufficient to produce desired results, given that the following GC form is used.

$$p(y_t|\mathbf{y_{t-1}}) = \sum_{l=1}^{L} d_l^{(t)} H_l(\frac{y_t}{s})\phi(\frac{y_t}{s})$$

where $s = \exp(\mu_h/2)$.

Since the volatility process in the eigen function approach of Meddahi (2001) uses a hermite approximation (HSV models), this framework also seems to be a plausible area to explore with the GC expansion.

The HSV model is given by:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \;\; \epsilon_t \sim N(0,1) \\ \sigma_t^2 &= a_0 + a_2(f_t^2 - 1) \\ f_t &= \beta f_{t-1} + \sqrt{1-\beta^2} v_t \;\; v_t \sim N(0,1) \end{aligned}$$

where $\theta = (a_0, a_2, \beta)$ are the parameters.

5.4 Using the orthogonality property

Several different approaches were followed to write the expansion in terms of hermite polynomials. However, we failed to devise an algorithm in this direction. An approach was to write $p(x_t|\mathbf{y_{t-1}})$ and $p(y_t|x_t)$ required in the recursion algorithm in GC form, and bypass the integration by using the orthogonality property of the hermite polynomials to compute $p(y_t|\mathbf{y_{t-1}})$. The main attraction is the use of the orthogonality property without resorting to the different numerical integration schemes. However, in order to implement this property, a part of the integrand ($p(y_t|x_t)$), with a hermite representation in terms of x_t takes the form of an extreme value distribution. Hence, the technique fails.

Another approach we tested was to write GC expansions for all pdfs. We carry out the following numerical procedure:

$$p_{x_t|\mathbf{y}_{t-1}}(x, \mathbf{y}_{t-1}) = \frac{\int p_{x_t|x_{t-1}}(x,x') \times p_{y_{t-1}|x_{t-1}}(y,x') \times p_{x_{t-1}|\mathbf{y}_{t-2}}(x', \mathbf{y}_{t-2}) dx'}{\int p_{y_{t-1}|x_{t-1}}(y,x') \times p_{x_{t-1}|\mathbf{y}_{t-2}}(x', \mathbf{y}_{t-2}) dx'}.$$

Writing $\phi_t(x)$ for $p_{x_t|\mathbf{y}_{t-1}}(x, \mathbf{y}_{t-1})$, suppressing the dependence on $\mathbf{y}_{t-1}$, gives

$$\phi_t(x) = \frac{\int A_y(x, x')\phi_{t-1}(x')dx'}{\int B_y(x')\phi_{t-1}(x')dx'}, \tag{5.1}$$

where

$$A_y(x, x') = p_{x_t|x_{t-1}}(x, x') \times p_{y_{t-1}|x_{t-1}}(y, x')$$

and

$$B_y(x') = p_{y_{t-1}|x_{t-1}}(y, x').$$

The initial condition is $\phi_0(x) = \phi(x)$.

Since the denominator in (5.1) is just a normalizing constant, it can also be written as

$$\psi_t(x) = \int A_y(x, x')\phi_{t-1}(x')dx'$$
$$\phi_t(x) = \frac{\psi_t(x)}{\int \psi_t(x)dx}.$$

With the GC representation

$$\phi_t(x) = \phi(x)\left[\sum_k \alpha_{t,k}H_k(x)\right],$$

an expansion

$$A_y(x, x') = \phi(x)\sum_l \sum_{l'} a_{l,l'}(y)H_l(x)H_{l'}(x')$$

would lead to

$$\psi_t(x) = \phi(x)\sum_l H_l(x)\sum_{l'} a_{l,l'}(y)\sum_k \alpha_{t-1,k}\int H_{l'}(x')H_k(x')\phi(x')dx'.$$

The integral is known, so the coefficient

$$\sum_{l'} a_{l,l'}(y)\sum_k \alpha_{t-1,k}\int H_{l'}(x')H_k(x')\phi(x')dx'$$

of $H_l(x)$ could in principle be calculated.

The coefficients $a_{l,l'}(y)$ in the expansion,

$$a_{l,l'}(y) = \frac{1}{m!n!}\int\int A_y(x, x')\phi(x')H_m(x)H_n(x')dx'dx,$$

could be calculated up front for a grid of ys by numerical integration, and interpolated in the recursion to the observed values y_{t-1}.

Further, in the HSV case, we can further simplify the problem analytically. Let us restrict the number of coefficients till 4 and using only the even ordered terms, we can write the above as follows:

$$\phi_t(x) = \phi(x)\left\{\alpha_{t,0}H_0(x) + \alpha_{t,2}H_2(x) + \alpha_{t,4}H_4(x)\right\}$$

If we start with GC expansions where $p_{x_t|x_{t-1}}(x,x') = \phi(x)\left\{\sum_i a_i(x')H_i(x)\right\}$ and $p_{y_{t-1}|x_{t-1}}(y,x') = \phi(y)\left\{\sum_j b_j(x')H_j(y)\right\}$ which in the HSV case, can be written as:

$$p_{y_{t-1}|x_{t-1}}(y,x') = \phi(y)\left\{\sum_j c_j(y)H_j(x')\right\}$$

Multiplying these two forms to obtain $A_y(x,x')$, we get

$$A_y(x,x') = \phi(x)\sum_l\sum_{l'} a_l(x')c_{l'}(y)\phi(y)H_l(x)H_{l'}(x')$$

An analytical expression for $a_{l,l'} = a_l(x')c_{l'}(y)\phi(y)$ can be obtained.

$$\begin{aligned}
\psi_t(x) &= \int A_y(x,x')\phi_{t-1}(x')dx' \\
&= \int A_y(x,x')\phi(x')\sum_k \alpha_{t-1,k}H_k(x')dx' \\
&= \phi(x)\phi(y)\sum_l H_l(x)\sum_k \alpha_{t-1,k}\sum_{l'} c_{l'}(y)\int a_l(x')H_{l'}(x')H_k(x')\phi(x')dx'
\end{aligned}$$

Since $p_{x_t|x_{t-1}}(x|x')$ is known, we can find the coefficients of its GC expansion.

$$a_l(x') = \frac{\beta^l}{l!}H_l(x')$$

where β is the persistence parameter. Using the above expression above, expanding and using the orthogonality result, we get,

$$\begin{aligned}
\psi_t(x) = \phi(x)\phi(y)\Big[&H_0(x)\Big(\alpha_{t-1,0}c_0(y) + 2!\alpha_{t-1,2}c_2(y) + 4!\alpha_{t-1,4}c_4(y)\Big) + \\
&+ H_2(x)\gamma^2\Big(\alpha_{t-1,0}c_2(y) + \alpha_{t-1,2}c_0(y) + 4\alpha_{t-1,2}c_2(y) + 12\alpha_{t-1,2}c_4(y) + \\
&+ 12\alpha_{t-1,4}c_2(y) + 96\alpha_{t-1,4}c_4(y)\Big) + H_4(x)\gamma^4\Big(\alpha_{t-1,0}c_4(y) + \alpha_{t-1,2}c_2(y) \\
&+ 8\alpha_{t-1,2}c_4(y) + \alpha_{t-1,4}c_0(y) + 8\alpha_{t-1,4}c_2(y) + 72\alpha_{t-1,4}c_4(y)\Big)\Big]
\end{aligned}$$

The denominator in $\phi_t(x)$ is the leading coefficient corresponding to the $H_0(x)$ term. Let us denote the coefficients of the above expression by coeff_k, then the denominator is $\text{coeff}_0\phi(\text{y})$. Since we have written the updated $\phi_t(x)$ coefficients with respect to coefficients in $\phi_{t-1}(x')$, we can compute the pdfs' recursively. Therefore,

$$\phi_t(x) = \phi(x)\Big[H_0(x)\frac{\text{coeff}_0}{\text{coeff}_0} + H_2(x)\frac{\text{coeff}_2}{\text{coeff}_0} + H_4(x)\frac{\text{coeff}_4}{\text{coeff}_0}\Big] \tag{5.2}$$

$$= \phi(x)\Big[\alpha_{t,0}H_0(x) + \alpha_{t,2}H_2(x) + \alpha_{t,4}H_4(x)\Big] \tag{5.3}$$

Both numerical and analytical results show that the algorithm fails to conform appropriately to the GC form. Using GC approximations in all cases make the framework too restrictive. A test to see whether the GC expansion fits the known probability $p_{x_t|x_{t-1}}(x|x')$ shows that the fit gets better as the persistence parameter β goes from 1 towards 0. However, the parameter of interest β is usually close to 1.

The issue with this formulation is its reduced flexibility. $\alpha_{t,0}$ is 1 by formulation. Hence the base density cannot adapt itself to the unimodal and bimodal distribution in the HSV case. When the observed return series is close to zero, $p(x_t|\mathbf{y_{t-1}})$ is unimodal and with extreme returns, the form becomes bimodal. The fixed base distribution acts as a hindrance in making the approximation adaptive to the changing shapes. In conclusion, two important criteria while considering approximating a distribution with the GC density approximation is that the pdf to be approximated has to be close enough to the base distribution or the leading term. If the probability in its tails decay at an exponential rate, the convergence rates with which the density approximation by hermite polynomials matches the true density get higher. This would eventually lead to a parsimonious use of hermite polynomials which would lead to a practical implementation. The fewer parameters or coefficients used, the better is the estimation. This scheme fails to flexibly switch between unimodal to bimodal in HSV case, and results in an extreme value density with unsuitable base density, for one of the integrands, in the SV case. Hence the orthogonality property cannot be used in the NF scheme.

5.5 Empirical Results

The parameter settings in the estimation of SV model parameters in earlier chapters are used. The first technique given in the Section 5.3 is implemented on these parameter settings to check the estimated densities in the NF algorithm. Since $P(y_t|\mathbf{y_{t-1}})$ estimates are used for SV model estimation and VaR calculation purposes, graphical comparisons of

$P(y_t|\mathbf{y_{t-1}})$ under the different parameter settings are provided. The Figure 5.1 shows the $P(y_t|\mathbf{y_{t-1}})$ values obtained by Watanabe's method and by applying the GC technique. By GC technique, updating is with respect to the coefficients of the Hermite polynomials. The number of coefficients chosen to represent $p(x_t|\mathbf{y_{t-1}})$ is 4 and 8 in the SV and HSV models respectively. This technique is described in the Section 5.3.

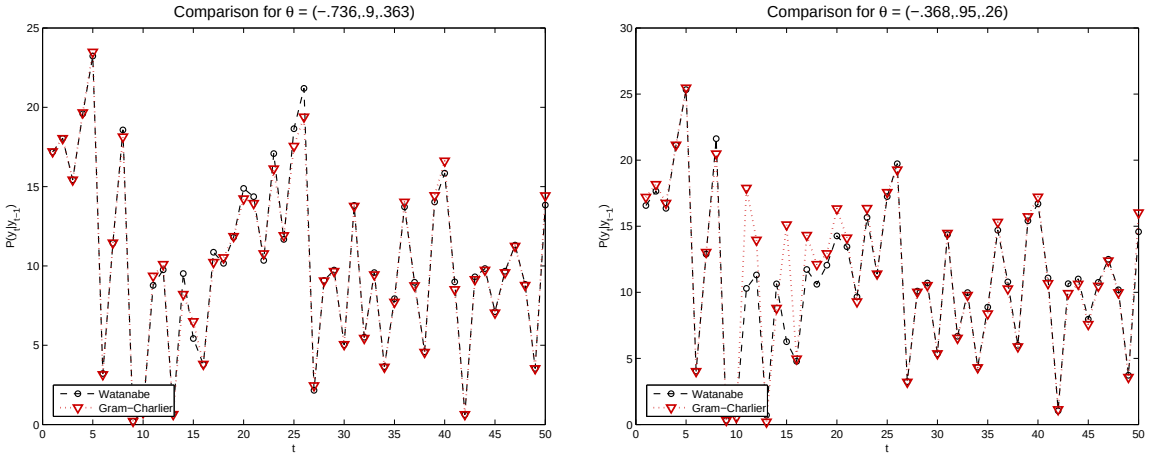

Figure 5.1: *Comparison of estimated conditional return probabilities calculated by Watanabe's method and that computed using the Gram-Charlier representation for θ=(-0.736,0.9,0.363) and θ=(-0.368,0.95,0.26)*

For the Hermite Stochastic volatility (HSV) model, 2000 data returns are generated with $\theta = (.7, .5, .98)$. The lower panel of Figure 5.2 shows a near perfect fit for the entire dataset. A smaller window of 50 datapoints from the set gives a better idea given in the upper panel of Figure 5.2.

As discussed earlier, in Finance, Cornish Fisher (CF) expansions are used to calculate the VaR of the data. For that purpose, GC coefficients are required for the $P(y_t|\mathbf{y_{t-1}})$ written in GC form. Hence, in such a case, a grid of the return values are decided apriori. For example, in the SV case, a range of [-0.3,0.3] is used with the number of coefficients chosen to be 8. This however, is an additional step in the aforementioned technique where we update coefficients of $P(x_t|\mathbf{y_{t-1}})$ terms only. This exercise yielded favorable results, and the plots are shown in Figure 5.3.

The 5% VaR plots are shown in Figure 5.4.

For estimation, 2000 data was simulated from a SV model and the results obtained using both the techniques are shown in Table 5.1. A single run study was conducted to

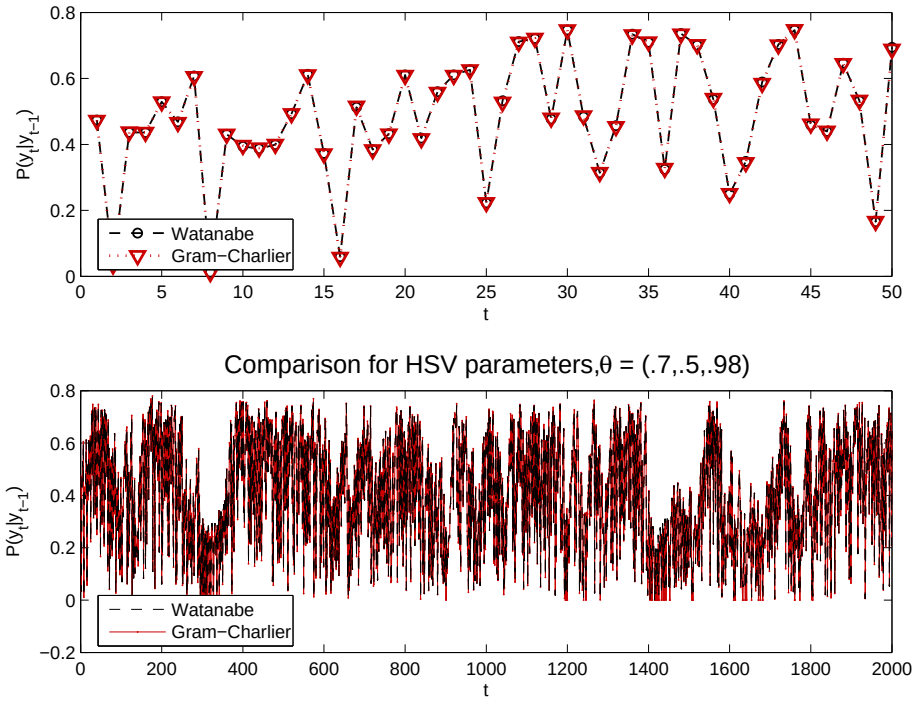

Figure 5.2: *Comparison of* $P(y_t|\mathbf{y_{t-1}})$ *calculated by Watanabe's method and that computed using the Gram-Charlier representation for* θ*=(0.7,0.5,0.98).*

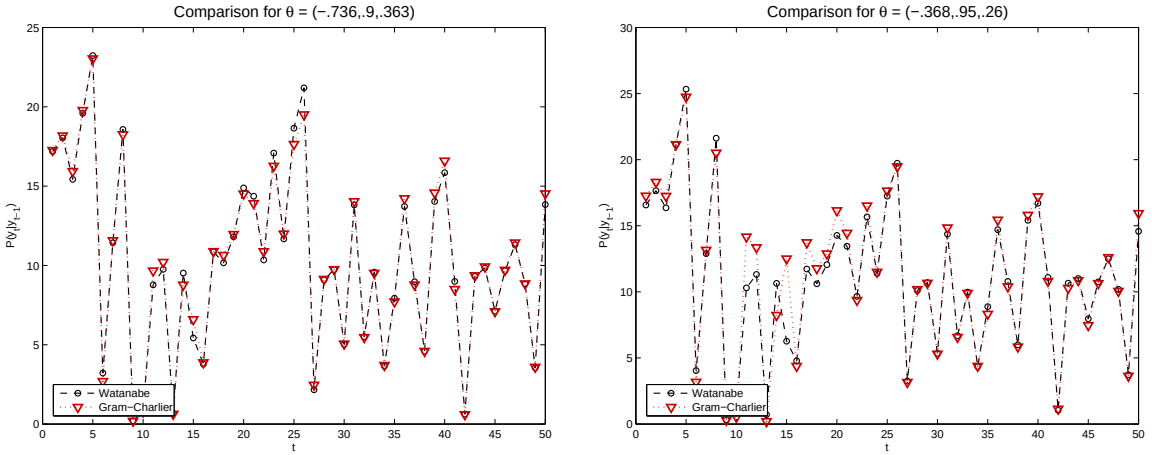

Figure 5.3: *Comparison of estimated conditional return probabilities calculated by Watanabe's method and that computed using the Gram-Charlier representation for CF application, where* $\theta = (-0.736, 0.9, 0.363)$ *and* $\theta = (-0.368, 0.95, 0.26)$

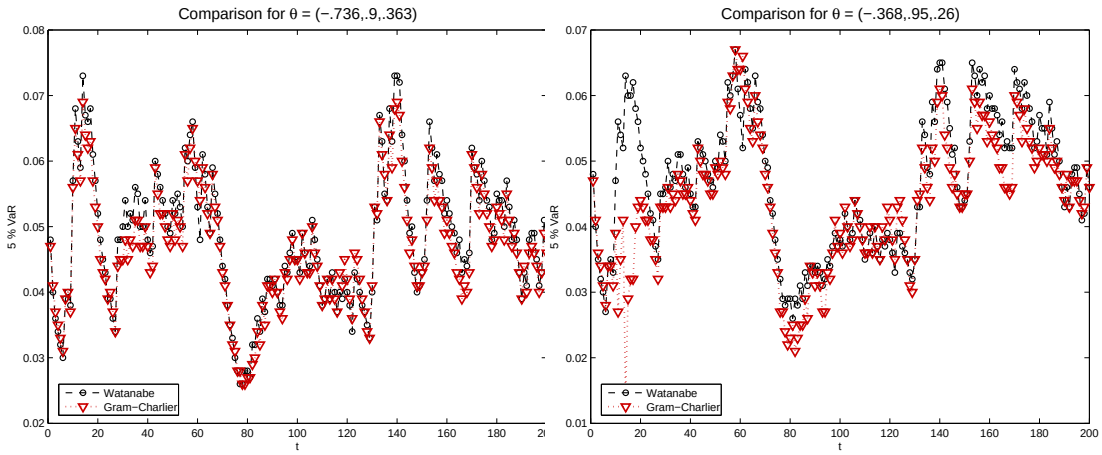

Figure 5.4: *Diagram showing the* 5% *VaR plots obtained from using the Watanabe's and GC approx in the NF scheme.*

compare the estimates from the Watanabe method with that of the Hermite approximation method. The set of parameter values considered are the same as above. The number of node points was chosen to be 50.

Table 5.1: *Estimates obtained by using Watanabe's and GC techniques in NFML*

True parameters	Hermite Approx.	Watanabe
	Estimates	Estimates
{-0.736, 0.9, 0.131769}	{-0.8547, 0.88, 0.1114 }	{-0.6878, 0.9072, 0.1537}
{-0.368, 0.95, 0.0676}	{-0.3598, 0.9544, 0.0635}	{-0.3722, 0.9497, 0.0682}
{-0.147, 0.98, 0.027556}	{-0.1489, 0.9847, 0.0271}	{-0.1529, 0.9795, 0.0284}

The above estimates were obtained by using constrained optimization techniques in matlab. The constraints used in the optimization is the constraints on the parameter space viz., $|\alpha_1| < 1$ and $\sigma_v^2 > 0$. In order to tackle the negative value calculation of the conditional probability values used in the computation of the loglikelihood, the minimum of the probability values are recorded and once a negative value is encountered, it is replaced with a uniformly distributed random number from $U(0, min)$ where min is the minimum value that was obtained till the last calculation. The hessian at the maximum likelihood estimate can be used to obtain the asymptotic standard errors of the estimates. Using the large sample theory results, we get,

$$\hat{\theta}_{MLE} \sim N(\theta_0, [I_T(\theta)]^{-1})$$

where $I_T(\theta)$ is the information matrix. The standard error is provided by its estimate given by the square root of $[I_T(\hat{\theta})]^{-1}$ which is the obtained by computing the hessian matrix at its MLE estimate ie, $\left[-\frac{\partial^2}{\partial\theta\partial\theta' LogL}\right]$.

5.6 Conclusions

Hermite polynomials are used in our analysis because of their orthogonality property and their support on the $\Re$ line. Further they can also be extended to Cornish Fisher type approximations to estimate quantiles. The Cornish Fisher implementation is left for future work.

The initial motivation of our endeavor relating to Hermite polynomials started because of the orthogonality property. SV models have been in the literature for a long time. However, computing the loglikelihood requires integration with respect to the latent state variables which are to be computed at every timestep, making the computation tedious. If a framework existed where the orthogonality property could be used, it would have eased the computations. However, from our extensive study, we could not obtain desirable results.

The updating scheme based on the coefficients of the Hermite expansions have shown promising results. Based on the first 4 coefficients in the SV case and 8 coefficients in the HSV case, we obtained results similar to the previous algorithm. The solution in each of the cases have been found to be comparable to the existing methods.

Following Watanabe using trapezoidal numerical integration and Fridman and Harris (1998) using Gauss Legendre numerical integration in SV models, Kawakatsu (2007) uses Gauss-Legendre approach on eigenfunction modeling of volatility process, the Hermite SV (HSV) models proposed by Meddahi (2001). The NFML scheme can also be extended to stochastic volatility models with leverage effects. Our scheme is applicable to all these areas.

The Gram Charlier type series is one way of obtaining an approximation to absolutely continuous distributions in terms of moments based on the Hermite polynomials. Gallant and Nychka (1987)'s representation of the semi-nonparametric (SNP) density can be modified to any other density having a moment generating function. The difference between SNP and GC is that the latter can take negative values, whereas the former is written as a square of the polynomial expansion. For ease of computation, GC is chosen over SNP which would otherwise lead to solving a quadratic equation at every step. Hence, the above can be extended to a SNP type density which might help in the optimization due

to its nonnegativity property.

Cornish fisher type of approximations can work really well when the distribution of the variable of interest has a closed form distribution. In the SV case, we are interested in finding the τ^{th} quantile of the current return given the past information, ie $y_t|\mathbf{y_{t-1}}$. Since the $P(y_t|\mathbf{y_{t-1}})$ does not have a closed form density, the direct implementation of Cornish Fisher is not possible. However, initial experiments with the SV case suggests that the $P(y_t|\mathbf{y_{t-1}})$ can be approximated very well in a Hermite expansion form. Hence, a Cornish Fisher type expansions as used by Jaschke and Jiang (2001) can be extended.

Chapter 6

Conclusion and Future Work

The objective of this dissertation was to propose a methodology that can be used to obtain parameter estimates of the SV model as well as the VaR estimates directly. The exact likelihood calculation of SV models in itself poses a challenging problem and many numerical simulation based procedures are available. Calculating the VaR in these scenarios pertains to a two step method. Robust methods are an growing research area. Most of the methods are concerned with model estimation and computing volatility forecasts used to filter the VaRs. Current research is focussed to develop tools that are better adapted to handling misspecification in models, data containing outliers, etc. Tools that yield robust statistics are especially important when dealing with real data applications eg. in finance, biostatistics, etc. Using QR as a tool to yield robust statistics and VaR calculation provides a solution to the problem. In this dissertation, we propose two methodologies Regression Quantile method of moments (RQMM) and Regression Quantile - Kalman Filtering (RQ-KF) scheme. In Chapters 3 and 4, we proposed the algorithms of the two models and illustrated the use of the two approaches by means of simulation studies (Sections 3.5 and 4.3.2) and VaR computation with these tools through real data application and simulation studies. We compared the two methods with established procedures: RQMM with EMM (both simulation based approaches) and RQ-KF with QML (both simplified approaches) for SV model parameter estimation. In the case of RQMM and EMM, comparison of the two approaches in high kurtosis data and model misspecification shows that RQMM is highly efficient in producing robust estimates in contrast to EMM. Also, in small sample studies with standard normal assumptions RQMM shows better estimates in terms of accuracy and efficiency based on MC bias and MSE calculations. In case of moderate samples RQMM

provides competitive results with respect to EMM in terms of efficiency when the distributional assumptions are correctly specified. RQ-KF results are also found to be competitive with respect to QML. Empirical findings show that these methods provide very good estimates especially in leptokurtic data. VaR evaluation results obtained using both RQMM and RQ-KF give substantial proof in their ability to produce reliable VaR estimates in terms of real data application and simulation studies.

The simulation studies are conducted with an emphasis on high kurtosis data motivated by financial data where leptokurtosis is an important and well documented data feature. Hence, preliminary results based on data analysis show the usefulness of the proposed methodologies with respect to applications in financial data.

In EMM estimators, finding scores under a distributional assumption of the error process with an appropriate auxiliary model is tantamount to considering an approach that results in a consistent and efficient estimate, given that the assumptions are correct. However, replacing the score function with the check function, the estimation procedure is rendered robust under outliers and misspecification. In real data scenarios where a greater possibility of nonstandard situations may arise, QR related tools are very useful. Hence the proposed approaches are relevant.

Following are some of the potential areas for future research.

RQMM is a flexible distribution free method than can be extended in several directions. RQMM belongs to the spectrum of Indirect Inference (II) methods. A simulation based II methods are flexible indispensable methods when the objective criterion of the structural model to be optimized has an intractible form. The objective criteria could be the likelihood function, moment type estimating equations as seen in Gourieroux *et al.* (1993) or could be the RQMM criteria as discussed in this study. Widely used examples of these techniques include discrete time stochastic volatility models, continuous time diffusion models, multinomial choice models, models with latent nonlinear effects, dynamic stochastic general equilibrum models, with applications to biostatistics as well. The application of RQMM in all these areas are an interesting future exercise.

RQ-KF is a simplified approach which finds its application in moderate to large sample sized data. A restriction of RQ-KF is that it can be extended only to models that can be written in a state space form where the Kalman filter is applicable. The advantage of using this method is that it is a computationally simple method to implement, and since relatively large datasets are available, this tool has its applications. Kalman filters can be

written for state space models with correlated errors; Harvey and Shephard (1996) showed, with some transformations, that the Asymmetric SV (ASV) models can be written in a state space form with uncorrelated errors. Hence, the algorithm used here can be easily extended.

The Robust Efficient Method of Moments (REMM) method introduced by Ortelli and Trojani (2005) proposes two algorithms which are designed to produce robust statistics by bounding the influence function corresponding to the EMM estimator. Comparing this method with RQMM when the underlying data is generated from the basic SV model and also other variations of SV is left for future work.

Since the auxiliary model selection in EMM with SNP is based on an adaptive algorithm based on AIC (Gallant and Tauchen, 1996) A similar extension in terms of the Cornish Fisher (CF) approximation to incorporate more general quantile specifications can be made for the auxiliary models in RQMM setup. $p(y_t|\mathbf{y_{t-1}})$ can be written in terms of a Hermite expansion. Let z_t denote the standardized version of the y_t's in this case. When a CF expansion for a z_t exists, the pdf of z_t can be represented as a Gram Charlier (GC) density, the conditional quantile can be deduced by multiplying with the scale. The scaling factor is accounted by the GARCH(1,1) or EGARCH models. An important goal in this direction would be to find an algorithm that uses the CF expansion and chooses the truncation point appropriate to the data under study. This is left for future work. Also, in Chapter 5, we have already shown that $p(y_t|\mathbf{y_{t-1}})$ can be represented in GC form. Finding a suitable CF expansion in terms of the truncation point for use in NF scheme is also left for future work.

The orthogonality property is a very attractive property in Hermite polynomials. The different algorithms considered in Chapter 5 were motivated to exploit this property. Finding an algorithm which uses this principle would significantly reduce the computational efforts required to calculate the exact likelihood in SV. This remains an interesting area for future research.

Bibliography

Andersen T., Chung H. and Sorensen B. (1999). Efficient method of moments estimation of a stochastic volatility model: A monte carlo study. *Journal of Econometrics* **91**, 61–87.

Andersen T. and Sørensen B. (1996). GMM estimation of a stochastic volatility model: a monte carlo study. *Journal of Business and Economic Statistics* **14**, 328–352.

Anderson B. and Moore J. (1979). *Optimal Filtering*. Prentice Hall.

Andrews D.W.K. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* **4**.

Angrist J., Chernozhukov V. and Fernandez-Val I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **74**, 539–563.

Asai M. (2008). Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models. *Journal of Empirical Finance* **15**, 332–341.

Bachelier L. (1900). *Theorie de la speculation*. Annales Scientifiques de l'Ecole Normale Superieure, Troisieme serie.

Bansal R., Gallant A., Hussey R. and Tauchen G. (1993). *Computational aspects of nonparametric simulation estimation*. Kluwer Academic Publishers, Boston, MA.

Bansal R., Gallant A., Hussey R. and Tauchen G. (1995). Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* **66**, 251–287.

Bollerslev T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Boudoukh J., Richardson M. and Whitelaw R. (1998). The best of both worlds. *Risk* **11**.

Chernozhukov V. (1999). Specification and other test processes for quantile regression. Technical report.

Christoffersen P. and Pelletier D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics* **2**, 84–108.

Cox J.C. and Ross S.A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics* **3**, 145–166.

Danielsson J. (1994). Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics* **64**, 375–400.

Danielsson J. and Richard J. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* **8**, S153–S173.

Duffie D. and Singleton K. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica* **61**, 929–952.

Engle R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**.

Engle R.F. and Manganelli S. (2004). CAViaR: Conditional Autoregressive Value at Risk by regression quantiles. *Journal of Business and Economic Statistics* **22**.

Fridman M. and Harris L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics* **16**, 284–291.

Gallant A. and Long J. (1997). Estimating stochastic differential equations efficiently by minimum chi- squared. *Biometrika* **84**, 125–141.

Gallant A. and Tauchen G. (1996). Which moments to match? *Econometric Theory* **12**, 657–681.

Gallant R. and Nychka D. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrics* **55**, 363–390.

Gnedenko B.V. (1943). Sur la distribution limite du terme maximum of d'une serie aleatorie. *Annals of Mathematics* **44**.

Gourieroux C., A. M. and Renault E. (1993). Indirect inference. *Journal of Applied Econometrics* **8**.

Gourieroux C. and Monfort A. (1996). *Simulation-Based Econometric Methods*. Oxford University Press, New York.

Hansen L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

Harvey A. (1989). *Forecasting, Structural Time Series and the Kalman filter*. Cambridge University Press, Cambridge.

Harvey A., Ruiz E. and Shephard N. (1993). Multivariate stochastic variance models. *Review of Economic studies* **61**.

Harvey A.C. and Shephard N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business Economic Statistics* **14**.

Heston S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**.

Huber P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* pp. 221–233.

Hull J. and White A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**.

Hull J. and White A. (1988). An analysis of the bias in option pricing caused by a stochastic volatility. *Advances in Futures and Options Research* **3**.

Jacquier E., Polson N. and Rossi P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**.

Jacquier E., Polson N. and Rossi P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* **122**.

Jaschke S. and Jiang Y. (2001). *Approximating Value at Risk in Conditional Gaussian Models*, volume 14.

Jenkinson A.F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meterological Society* **81**.

Johnson H. and Shanno D. (1987). Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis* **22**.

Jorion P. (2007). *Value at Risk the New Benchmark for Managing Financial Risk*. McGRaw-Hill.

Kawakatsu H. (2007). Numerical integration-based Gaussian mixture filters for maximum likelihood estimation of asymmetric stochastic volatility models. *Econometrics* **10**, 342–358.

Kim S., Shephard N. and Chib S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**, 361–393.

Kim T.H. and White H. (2002). Estimation, inference, and specification testing for possibly misspecified quantile regressions. *Advances in Econometrics* **17**, 107–132.

Kitagawa G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**, 1032–1041.

Koenker R.W. and Bassett G. (1978). Regression quantiles. *Econometrica* **46**.

Lee Y. and Lee M. (1992). On the derivation and computation of the Cornish-Fisher expansion. *Australian and New Zealand Journal of Statistics* **34**, 443–450.

Liesenfeld R. and Jung R.C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* **15**, 137–160.

Logerstaey J. and More L. (1995). *Introduction to Riskmetrics*. Morgan Guarantee Trust Company, New York.

McFadden D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57**, 995–1026.

Meddahi N. (2001). An eigenfunction approach for volatility modeling. *Unpublished* pp. 1–44.

Melino A. and Turnbull S. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* **45**.

Nelson D. (1988). *The Time Series behaviour of stock market volatility and returns*. Ph.D. thesis.

Nelson D. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* **59**.

Ortelli C. and Trojani F. (2005). Robust efficient method of moments. *Journal of Econometrics* **128**, 69–97.

Osborne M. (1959). Brownian motion in the stock market. *Operations Research* **7**.

Pakes A. and Pollard D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

Powell J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25**, 303–325.

Powell J. (1991). *Estimation of Monotonic Regression Models Under Quantile Restriction,in: Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press.

Powell J.L. (1986). Censored regression quantiles. *Journal of Econometrics* **32**, 143–155.

Ruiz E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* **63**.

Sandmann G. and Koopman S. (1998). Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics* **87**, 271–301.

Schwert G.W. (1989). Why does stock market volatility change over time. *Journal of Finance* **44**, 1115–1153.

Scott L.O. (1987). Option pricing when the variance changes randomly: theory, estimation and an application. *Journal of Financial and Quantitative Analysis* **22**.

Shephard N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

Stein E.M. and Stein J.C. (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies* **4**.

Tanizaki H. (1996). *Nonlinear Filters:Estimation and Applications*. Springer-Verlag, Berlin.

Taylor S. (1986). *Modeling Financial Time Series*. John Wiley and Sons.

Tsay R.S. (1987). Conditional heteroscedastic time series models. *Journal of American Statistical Association* **82**.

Watanabe T. (1999). A non-linear filtering approach to stochastic volatility models with an application to daily stock returns. *Journal of Applied Econometrics* **14**.

Watanabe T. and Asai M. (2003). Stochastic volatility models with heavy-tailed distributions: a bayesian analysis. *Unpublished* pp. 1–44.

Weiss A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* **7**, 46–68.

White H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

White H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York.

White H. and Domowitz I. (1984). Nonlinear regression with dependent observations. *Econometrica* **52**.

Wiggins J.B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* **19**.

Yu J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* **127**, 165–178.

Appendices

Appendix A

Technical Details

Proof of Theorem 3.3.1. The proof follows by showing that finding the argument minimum of the check function with respect to the auxiliary parameter is equivalent to minimizing the distance between the check functions of the true parameters from the proxy since the minimization is with respect to β. Rearranging in terms of Δ_τ and ϵ_τ,

$$
\begin{aligned}
& E_\theta[(\tau - I(y_t < q_t(\beta)))(y_t - q_t(\beta))] \\
& \quad -E_\theta[(\tau - I(y_t < Q_{t,\tau}(y_t|X_t)))(y_t - Q_{t,\tau}(y_t|X_t))] \\
= \; & E_\theta\Big[\tau(Q_\tau(y_t|X_t) - q_\tau(\beta)) - I(y_t < q_t(\beta))(y_t - q_t(\beta)) \\
& \quad +I(y_t < Q_{t,\tau}(y_t|X_t))(y_t - Q_{t,\tau}(y_t|X_t))\Big] \\
= \; & E_\theta\Big[\tau(Q_\tau(y_t|X_t) - q_\tau(\beta)) - I(\epsilon_{\tau,t} < \Delta_\tau(X,\beta))(\epsilon_{\tau,t} - \Delta_\tau(X,\beta)) \\
& \quad +I(\epsilon_{\tau,t} < 0)(\epsilon_{\tau,t})\Big] \\
= \; & E_\theta\Big[(-\tau(\Delta_\tau(X,\beta)) - I(\epsilon_{\tau,t} < \Delta_\tau(X,\beta))(\epsilon_{\tau,t} - \Delta_\tau(X,\beta)) \\
& \quad +I(\epsilon_{\tau,t} < 0)(\epsilon_{\tau,t})\Big] \\
= \; & E_\theta\Big[(\tau - I(\epsilon_{\tau,t} < \Delta_\tau(X,\beta)))(\epsilon_{\tau,t} - \Delta_\tau(X,\beta)) \\
& \quad -(\tau - I(\epsilon_{\tau,t} < 0))\epsilon_{\tau,t}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&= E_\theta\{E[(I(\epsilon_{\tau,t} < \Delta_\tau(X,\beta)) - \tau)\Delta_\tau(X,\beta)|X]\} \\
&\quad -E_\theta\{E[\epsilon_{\tau,t}(I(\epsilon_{\tau,t} < \Delta_\tau(X,\beta)) - I(\epsilon_{\tau,t} < 0))]\} \\
&= E_\theta(F_{\epsilon_{\tau,t}}(\Delta_\tau(X,\beta)|X) - F_{\epsilon_{\tau,t}}(0|X)\Delta_\tau(X,\beta)) \\
&\quad -E_\theta(I(\epsilon_{\tau,t} \in [0, \Delta_\tau(X,\beta)])\epsilon_{\tau,t}|X) \\
&= E_\theta \int_0^1 (1-u) f_y(uq_\tau(\beta) + (1-u)Q_\tau(y_t|X)|X) du \Delta_\tau^2(X,\beta)
\end{aligned}
$$

The final result is obtained by taking the transformation, $u = \epsilon/\Delta(X,\beta)$.

Consistency assumptions for Theorem 3.4.1 (White (1994), Kim and White (2002)) The following assumptions are needed to guarantee the consistency of $\hat{\beta}$.

CB1 (Ω, F, P) is a complete probability space and $\{\epsilon_t, x_t\}, t = 1, 2, \ldots,$ are random vectors on this space.

CB2 The function $q_t(\beta) : \Re^k \times B \to \Re$ is such that for each β_τ in B, a compact subset of $\Re^q$, $q(x_t, \beta_\tau)$ is measurable with respect to the Borel set B^k and $q_t(\beta)$ is continuous in B, a.s. in Prob, t $= 1, 2, \ldots$ for a given choice of explanatory variables.

CB3 There exists a $\delta > 1$, such that $E|\nabla q_t(\beta)|^\delta < \infty$.

CB4
- $E([I(y_t < q_t(x_t, \beta_\tau)) - \tau][y_t - q_t(x_t, \beta_\tau)])$ exists and is finite for each β_τ in B.
- $E([I(y_t < q_t(x_t, \beta_\tau)) - \tau][y_t - q_t(x_t, \beta_\tau)])$ is continuous in β_τ.
- $([I(y_t < q_t(x_t, \beta_\tau)) - \tau][y_t - q_t(x_t, \beta_\tau)])$ obeys the strong (weak) law of large numbers. We could assume that ϵ_t, x_t are $\alpha-$mixing, i.e. $\alpha(m) \to 0$ as $m \to \infty$. See Andrews (1988) and White and Domowitz (1984), for details.

CB5 $1/TE[I(y_t < q_t(x_t, \beta)) - \tau][y_t - q_t(x_t, \beta)]$ has identifiably unique maximisers.

CB6 For all x, $f_{\epsilon|x}(0|x) > 0$

Asymptotically Normality assumptions for Theorem 3.4.2 (Kim and White (2002)) The following conditions are necessary for the asymptotic normality result.

ANB1 $\nabla_t q_t(\beta_\tau)$ is A-smooth with variables A_{it} and functions $\rho_i, i = 1 \ldots q$. In addition, $max_i \rho_i(d) \leq d$ for $d > 0$ small enough. [1].

[1] $q_t(\beta)$ is A-smooth with variables $A_{0,t}$ and function ρ if, for each $\beta \in B$, there is a constant $\tau > 0$ such that $||\beta^* - \beta|| \leq \tau$ implies that $|q(x_t, \beta^*) - q(x_t, \beta)| \leq A_{0t}(x_t)\rho(||\beta^* - \beta||)$ for all t, a.s. - Prob, where A_{0t} and ρ are nonrandom functions such that $A_{0t}(x_t)$ is a random variable, $\limsup_{T\to\infty} T^{-1} \sum E[A_{0t}(x_t)] < \infty$, $\rho(\nu) > 0$ for $\nu > 0, \rho(\nu) \to 0$ as $\nu \to 0$ and τ, A_{0t}, ρ and the null set may depend on β

ANB2
- $f_t(\epsilon)$ is Lipschitz continuous in ϵ uniformly in t.
- For each t and (ϵ, ν), $f_t(\epsilon, \phi, \nu)$ is continuous in ϕ and bounded.

ANB3 For each t and s, $u_t(\phi, \beta_\tau, s)$ is continuous in (ϕ, β_τ).

ANB4 $\{\epsilon_t, X_t\}$ are α-mixing, with parameter $\alpha(m)$, and there exist $\Delta < \infty$ and $r > 2$ such that $\alpha(m) < \Delta m^\lambda$ for some $\lambda < -2r/(r-2)$.

ANB5 For some $r > 2$, $\nabla_i q_t(\beta)$ is uniformly r-dominated by functions a_{1t}.

ANB6 For all t and i, $E|\sup_\beta \mathrm{A_{it}}|^\mathrm{r} < \Delta_1 < \infty$. There exist measurable functions a_{2t} such that $|u_t| < a_{2t}$ and for all t, $\int a_{2t} d\nu < \infty$ and $\int (a_{1t})^3 a_{2t} d\nu < \infty$.

ANB7 There exists a matrix A such that $T^{-1} \sum_{t=a+1}^{a+T} E[\nabla q_t(\beta) \nabla' q_t(\beta)] \to A$ as $T \to \infty$, uniformly in a.

ANB8 There exists a matrix $D = E[f_{\epsilon|X}(0|X) \nabla q(\beta) \nabla q(\beta)']$ positive definite.

ANB9 There exists a positive definite matrix $A_0 = E[(\tau - I(y_t < q_t(\beta)))(\tau - I(y_t < q_t(\beta)))' \nabla q(\beta) \nabla q(\beta)']$

Consistency assumptions for Theorem 3.4.3 (following Gourieroux and Monfort (1996)) To prove consistency, the following conditions are necessary ((Gourieroux and Monfort, 1996)).

T1 The $RQ_T(\tilde{y}_T; \beta)$ tends almost surely to a deterministic limit function $RQ_\infty(\theta, \beta)$ uniformly in (θ, β) when $T \to \infty$. $\tilde{Y}$ denotes the simulated y series in the second step.

T2 The limit function has a unique maximum with respect to β: $b(\theta) = \mathrm{argmin}_\beta \, RQ_\infty(\theta, \beta)$.

T3 RQ_T and RQ_∞ are differentiable with respect to β, and $\dfrac{\partial RQ_\infty(\theta, \beta)}{\partial \beta} = \lim_T \frac{\partial RQ_T(\tilde{y}_T(\theta); \beta)}{\partial \beta}$

T4 The only solution of the asymptotic first order conditions is $b(\theta) : \dfrac{\partial RQ_\infty(\theta, \beta)}{\partial \beta} = 0 \Rightarrow \beta = b(\theta)$.

T5 The equation $\beta = b(\theta)$ admits an unique solution.

The proof of the consistency is that since both estimators converge to the binding function in the limit and one converges to $b(\theta_0)$, the estimator can only converge to θ (using (T4) and (T5)), since the solution to the binding function is unique.

Asymptotic Normality proof for Theorem 3.4.3 The RQ criterion being an M-type estimator, the proof follows directly from Gourieroux and Monfort (1996). Additional regularity conditions concerning the second order differentiability of the RQ criterion function with respect to both parameters and continuity of the derivatives are required. We assume that:

$$J_0 = \operatorname*{plim}_T \frac{-\partial^2 RQ_T}{\partial\beta\partial\beta'}(y_T, X_T; b(\theta_0)),$$
$$\sqrt{T}\frac{\partial RQ_T(y_T, b(\theta_0))}{\partial\beta} \xrightarrow{D} N(0, S_0)$$

We consider the asymptotic expansion of $\hat{\beta}$ estimators. We start with:

$$\sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0)), \hat{\beta}_N(\theta_0))}{\partial\beta} = 0$$

Expanding around the limiting value θ_0, $b(\theta_0)$,

$$\sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta} + \frac{\partial^2 RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta\partial\beta'}\sqrt{T}[\hat{\beta}_N(\theta_0) - b(\theta_0)] = o_p(1),$$
$$\sqrt{T}[\hat{\beta}_N(\theta_0) - b(\theta_0)] = \left[-\frac{\partial^2 RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta\partial\beta'}\right]^{-1} \times \sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta} + o_p(1)$$
$$= \left[-\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\beta\partial\beta'}\right]^{-1} \times \sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta} + o_p(1)$$
$$\sqrt{T}[\hat{\beta}_N(\theta_0) - b(\theta_0)] = J_0^{-1}\sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta} + o_p(1)$$

Again starting from the minimization problem, the first order condition is:

$$\left[\frac{\partial^2 RQ_T(\tilde{Y}(\hat{\theta}), \hat{\beta}_T)}{\partial\theta\partial\beta'}\right]\Sigma\left[\frac{\partial RQ_T(\tilde{Y}(\hat{\theta}), \hat{\beta}_T)}{\partial\beta}\right] = 0$$

An expansion around θ_0, $b(\theta_0)$ gives:

$$\left[\frac{\partial^2 RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\theta\partial\beta'}\right]\Sigma\left[\sqrt{T}\frac{\partial RQ_T(\tilde{Y}(\theta_0), b(\theta_0))}{\partial\beta} + \right.$$
$$\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\beta\partial\beta'}\sqrt{T}[\hat{\beta}_T - b(\theta_0)] +$$
$$\left.\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\beta\partial\theta'}\sqrt{T}[\hat{\theta}(\Sigma) - \theta_0]\right] = o_p(1)$$

Using the previous relation,

$$\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\theta\partial\beta'}\Sigma\left[J_0\sqrt{T}[\hat{\beta}_N(\theta_0) - \hat{\beta}_T] + \frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\beta\partial\theta'}\sqrt{T}(\hat{\theta}(\Sigma) - \theta_0)\right] = o_p(1)$$

Hence, we get:

$$\sqrt{T}(\hat{\theta}(\Sigma) - \theta_0) = \left[\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\theta\partial\beta'}\Sigma\frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\beta\partial\theta'}\right]^{-1} \times \frac{\partial^2 RQ_\infty(\theta_0, b(\theta_0))}{\partial\theta\partial\beta'}\Sigma J_0\sqrt{T}[\hat{\beta}_T - \hat{\beta}_N(\theta_0)] + o_p(1)$$

Combining the asymptotic normality result above, we get the asymptotically normality result for $\hat{\theta}(\Sigma)$.

Analytical derivation of gradient of the log likelihood

$$\begin{aligned} L &= \prod_{t=1}^{T} p_\theta(y_t|\mathbf{y_{t-1}}) \\ \log L &= \sum_{t=1}^{T} \log p_\theta(y_t|\mathbf{y_{t-1}}) \end{aligned}$$

The gradient calculation is to be computed term by term with a recursive scheme and $\theta = \{\alpha_0, \alpha_1, \sigma_v^2\}$. The gradient for $t = 1$ is derived, followed by the other terms of the recursion which is based on the first gradient. The $\log p(y_t|\mathbf{y_{t-1}})$ when $t = 1$ involves $p(x_t|\mathbf{y_{t-1}})$ which has been taken to be the unconditional distribution of $p(x_1) \equiv N_{x_1}(\mu_h, \sigma_h^2)$, where $\mu_h = \frac{\alpha_0}{(1-\alpha_1)}$ and $\sigma_h^2 = \frac{\sigma_v^2}{(1-\alpha_1^2)}$. The gradient takes this into account:

$$\frac{\partial logL(\theta)}{\partial\theta} = \frac{\partial \log p_\theta(y_1|y_0)}{\partial\theta} + \frac{\partial \log p_\theta(y_2|\mathbf{y_1})}{\partial\theta} + \ldots + \frac{\partial \log p_\theta(y_T|\mathbf{y_{T-1}})}{\partial\theta}$$

where

$$\frac{\partial \log p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta} = \frac{1}{p_\theta(y_t|\mathbf{y_{t-1}})}\frac{\partial p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta}$$

The $p_\theta(.)$ term denotes that the pdf is θ-dependent. Since, the initial integration term assumes $p_\theta(x_t|\mathbf{y_{t-1}})$ to be the unconditional distribution, the derivation of the gradient for the first term would be different from the rest. However, after the first term, the remaining terms would have the same form, hence only the derivation for the second time series point will suffice.

$$\begin{aligned} \frac{\partial p_\theta(y_1|y_0)}{\partial\theta} &= \frac{1}{\partial\theta}\int p(y_1|x_1)p_\theta(x_1|y_0)dx_1 \\ &= \frac{1}{\partial\theta}\int p(y_1|x_1)p_\theta(x_1)dx_1 \end{aligned}$$

Given that $p(y_t|x_t) \sim N(0, e^{x_t})$, $p(x_t) \sim N(\mu_h, \sigma_h^2)$ and $p(x_t|x_{t-1}) \sim N(\alpha_0 + \alpha_1 x_{t-1}, \sigma_v^2)$.

$$\frac{\partial p_\theta(y_1|\mathbf{y_0})}{\partial \alpha_0} = \int N_{y_1}(0, e^{x_t}) \frac{1}{\sqrt{2\pi\sigma_h^2}} \frac{\partial}{\partial \alpha_0} e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_h^2}^2} dx_1$$

$$\begin{aligned}
\frac{\partial}{\partial \alpha_0} e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_h^2}^2} &= e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_h^2}^2} \left(\frac{\left(x_1 - \frac{\alpha_0}{1-\alpha_1}\right)}{\sigma_h} \right) \frac{1/(1-\alpha_1)}{\sigma_h} \\
\Rightarrow \frac{\partial p_\theta(y_1|\mathbf{y_0})}{\partial \alpha_0} &= \int N_{y_1}(0, e^{x_t}) N_{x_1}(\mu_h, \sigma_h) \left(\frac{x_1 - \mu_h}{\sigma_h} \right) \frac{1}{\sigma_h(1-\alpha_1)} dx_1
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \alpha_1} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_v^2/(1-\alpha_1^2)}} \times e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_h^2}^2} &= e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_h^2}^2} \frac{1}{\sigma_v\sqrt{2\pi}} \frac{\partial}{\partial \alpha_1} (\sqrt{1-\alpha_1^2}) + \\
&\quad + \frac{1}{\sigma_h\sqrt{2\pi}} \frac{\partial}{\partial \alpha_1} \left[e^{-\frac{1}{2}\frac{x_1 - \frac{\alpha_0}{1-\alpha_1}}{\sigma_v^2/(1-\alpha_1^2)}^2} \right] \\
&= (I) + (II)
\end{aligned}$$

where

$$\begin{aligned}
(I) &= \frac{1}{\sigma_v\sqrt{2\pi}} \frac{-\alpha_1}{\sqrt{1-\alpha_1^2}} e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}}, \\
(II) &= \frac{1}{\sigma_h\sqrt{2\pi}} \frac{\partial}{\partial \alpha_1} \left[e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} \right]
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{1}{\sigma_h\sqrt{2\pi}} \frac{\partial}{\partial \alpha_1} \left[e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} \right] &= \frac{1}{\sigma_h\sqrt{2\pi}} \left[e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} \times (-1/2) \right] \times \\
&\quad \times \left\{ \left(\frac{1-\alpha_1^2}{\sigma_v^2} \right) \frac{\partial}{\partial \alpha_1} \left(x_1 - \frac{\alpha_0}{1-\alpha_1} \right)^2 + \right. \\
&\quad \left. + \left(x_1 - \frac{\alpha_0}{1-\alpha_1} \right)^2 \frac{\partial}{\partial \alpha_1} \left(\frac{1-\alpha_1^2}{\sigma_v^2} \right) \right\} \\
&= \frac{1}{\sigma_h\sqrt{2\pi}} \left[e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} \times (-1/2) \right] \times \\
&\quad \times \left\{ \left(\frac{1-\alpha_1^2}{\sigma_v^2} \right) \frac{\partial}{\partial \alpha_1} \left(\frac{-\alpha_0}{1-\alpha_1} \right) 2 \left(x_1 - \frac{\alpha_0}{1-\alpha_1} \right) + \right.
\end{aligned}$$

$$
\begin{aligned}
&+\left(x_1-\frac{\alpha_0}{1-\alpha_1}\right)^2\left(\frac{-2\alpha_1}{\sigma_v^2}\right)\Bigg\}\\
=\;& N_{x_1}(\mu_h,\sigma_h^2)\Bigg\{\left(\frac{1-\alpha_1^2}{\sigma_v^2}\left(x_1-\frac{\alpha_0}{1-\alpha_1}\right)\right)\left(\frac{\alpha_0}{(1-\alpha_1)^2}\right)\\
&+\left(x_1-\frac{\alpha_0}{1-\alpha_1}\right)^2\left(\frac{\alpha_1}{\sigma_v^2}\right)\Bigg\}\Bigg]
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\partial p(y_1|y_0)}{\partial\alpha_1} =\;& \int N_{y_1}(0,e^{x_1})\times\\
&\times\Bigg[\Bigg\{e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}}\left(\frac{-\alpha_1}{(1-\alpha_1^2)}\right)\left(\frac{1}{\sigma_h\sqrt{2\pi}}\right)\Bigg\}+\\
&+\Bigg\{N_{x_1}(\mu_h,\sigma_h^2)\left\{\left(\frac{1-\alpha_1^2}{\sigma_v^2}\right)(x_1-\mu_h)\frac{\alpha_0}{(1-\alpha_1)^2}+(x_1-\mu_h)^2\frac{\alpha_1}{\sigma_v^2}\right\}\Bigg\}\Bigg]dx_1\\
=\;& \int N_{y_1}(0,e^{x_1})\Bigg[\left(\frac{1}{\sigma_h\sqrt{2\pi}}\right)e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}}\times\\
&\times\Bigg\{\left(\frac{-\alpha_1}{(1-\alpha_1^2)}\right)+\left[\left(\frac{1-\alpha_1^2}{\sigma_v^2}\right)(x_1-\mu_h)\frac{\alpha_0}{(1-\alpha_1)^2}+(x_1-\mu_h)^2\frac{\alpha_1}{\sigma_v^2}\right]\Bigg\}\Bigg]dx_1
\end{aligned}
$$

This can be further simplified as follows:

$$
\begin{aligned}
&\left[\left(\frac{1-\alpha_1^2}{\sigma_v^2}\right)(x_1-\mu_h)\frac{\alpha_0}{(1-\alpha_1)^2}+(x_1-\mu_h)^2\frac{\alpha_1}{\sigma_v^2}\right]\\
=\;&\left[(x_1-\mu_h)\left[\left(\frac{(1-\alpha_1^2)\alpha_0}{\sigma_v^2(1-\alpha_1)^2}\right)+(x_1-\mu_h)\frac{\alpha_1}{\sigma_v^2}\right]\right]\\
=\;&\left[(x_1-\mu_h)\left[\frac{(1-\alpha_1^2)\alpha_0+\left(x_1(1-\alpha_1)^2-\alpha_0(1-\alpha_1)\right)\alpha_1}{\sigma_v^2(1-\alpha_1)^2}\right]\right]\\
=\;&\left[(x_1-\mu_h)\left[\frac{\alpha_0-\alpha_0\alpha_1^2+\alpha_1x_1(1-\alpha_1)^2-\alpha_1\alpha_0+\alpha_0\alpha_1^2}{\sigma_v^2(1-\alpha_1)^2}\right]\right]\\
=\;&\left[(x_1-\mu_h)\left[\frac{\alpha_0+\alpha_1x_1(1-\alpha_1)^2-\alpha_1\alpha_0}{\sigma_v^2(1-\alpha_1)^2}\right]\right]\\
=\;&\left[(x_1-\mu_h)\left[\frac{\alpha_0(1-\alpha_1)+\alpha_1x_1(1-\alpha_1)^2}{\sigma_v^2(1-\alpha_1)^2}\right]\right]\\
=\;&\left[(x_1-\mu_h)\left[\frac{\alpha_0+\alpha_1x_1(1-\alpha_1)}{\sigma_v^2(1-\alpha_1)}\right]\right]
\end{aligned}
$$

$$
\frac{\partial p(y_1|y_0)}{\partial\alpha_1}=\int N_{y_1}(0,e^{x_1})\Big[N_{x_1}(\mu_h,\sigma_h^2)\Big\{\Big(\frac{-\alpha_1}{(1-\alpha_1^2)}\Big)+\Big(x_1-\mu_h\Big)\Big[\frac{\alpha_0+\alpha_1x_1(1-\alpha_1)}{\sigma_v^2(1-\alpha_1)}\Big]\Big\}\Big]dx_1
$$

For notational convenience, let $\sigma_v^2 = \gamma$. Then

$$\begin{aligned}
\frac{\partial p(y_1|y_0)}{\partial \gamma} &= \int N_{y_1}(0, e^{x_1}) \times \\
&\quad \times \left[\left(\frac{1}{\sigma_h \sqrt{2\pi}} \right) \frac{-1}{2\gamma} e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} + \right. \\
&\quad \left. + \left(\frac{1}{\sigma_h \sqrt{2\pi}} \right) e^{-\frac{1}{2}\frac{(x_1-\mu_h)^2}{\sigma_h^2}} \frac{1}{2\gamma^2}(x_1-\mu_h)^2(1-\alpha_1)^2 \right] dx_1 \\
&= \int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \left\{ \frac{-1}{2\gamma} + \frac{1}{2\gamma^2}(x_1-\mu_h)^2(1-\alpha_1)^2 \right\} dx_1,
\end{aligned}$$

$$\frac{\partial \log p(y_2|y_1)}{\partial \theta} = \frac{1}{p(y_2|y_1)} \frac{\partial p(y_2|y_1)}{\partial \theta}$$

where

$$\frac{\partial p(y_2|y_1)}{\partial \theta} = \frac{\partial}{\partial \theta} \int p(y_2|x_2) p_\theta(x_2|y_1) dx_2,$$

$$\begin{aligned}
p_\theta(x_2|y_1) &= \int p_\theta(x_2|x_1) p_\theta(x_1|y_1) dx_1 \\
&= \int \frac{p_\theta(x_2|x_1) p(y_1|x_1) p_\theta(x_1) dx_1}{p_\theta(y_1|y_0)}
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial}{\partial \theta} \left[\int \frac{p_\theta(x_2|x_1) p(y_1|x_1) p_\theta(x_1) dx_1}{p_\theta(y_1|y_0)} \right] &= \frac{\partial}{\partial \theta} \left[\frac{A(\theta)}{B(\theta)} \right] \\
&= \frac{B(\theta)\frac{\partial}{\partial \theta} A(\theta) - A(\theta)\frac{\partial}{\partial \theta} B(\theta)}{B(\theta)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \theta} A(\theta) &= \frac{\partial}{\partial \theta} \int p_\theta(x_2|x_1) p(y_1|x_1) p_\theta(x_1) dx_1 \\
&= \int p(y_1|x_1) \left[p_\theta(x_1) \frac{\partial}{\partial \theta} p_\theta(x_2|x_1) + p_\theta(x_2|x_1) \frac{\partial}{\partial \theta} p_\theta(x_1) dx_1 \right]
\end{aligned}$$

The gradient of $p_\theta(x_2|x_1)$ is to be computed for all the parameters, and the gradients of $p_\theta(x_1)$ has already been derived above. Hence, the gradient of $p_\theta(x_2|x_1)$ with respect to α_0 is derived below.

$$\begin{aligned}
\frac{\partial}{\partial \alpha_0} p_\theta(x_2|x_1) &= \frac{\partial}{\partial \alpha_0} \left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}} \right) e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1 x_1)^2}{\gamma}} \\
&= \left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}} \right) e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1 x_1)^2}{\gamma}} \times (-1/2) 2 \left(\frac{x_2-\alpha_0-\alpha_1 x_1}{\sqrt{\gamma}} \right) \left(\frac{-1}{\sqrt{\gamma}} \right) \\
&= \left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}} \right) e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1 x_1)^2}{\gamma}} \times \left(\frac{x_2-\alpha_0-\alpha_1 x_1}{\gamma} \right)
\end{aligned}$$

The gradient of $p_\theta(x_2|x_1)$ with respect to α_1 is as follows.

$$\begin{aligned}\frac{\partial}{\partial\alpha_1}p_\theta(x_2|x_1) &= \frac{\partial}{\partial\alpha_1}\left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}}\right)e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{\gamma}} \\ &= \left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}}\right)e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{\gamma}}\times\left(\frac{x_2-\alpha_0-\alpha_1x_1}{\sqrt{\gamma}}\right)\left(\frac{x_1}{\sqrt{\gamma}}\right)\end{aligned}$$

The gradient of $p_\theta(x_2|x_1)$ with respect to $\sigma_v^2=\gamma$ is as follows.

$$\begin{aligned}\frac{\partial}{\partial\gamma}p_\theta(x_2|x_1) &= \frac{\partial}{\partial\gamma}\left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}}\right)e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{\gamma}} \\ &= \left(\frac{1}{\sqrt{\gamma}\sqrt{2\pi}}\right)e^{-\frac{1}{2}\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{\gamma}}\times\left(\frac{-1}{\gamma}+\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{2\gamma^2}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\alpha_0}p_\theta(x_1) &= N_{x_1}(\mu_h,\sigma_h)\left(\frac{x_1-\mu_h}{\sigma_h}\right)\frac{1}{(1-\alpha_1)\sigma_h}, \\ \frac{\partial}{\partial\alpha_1}p_\theta(x_1) &= N_{x_1}(\mu_h,\sigma_h)\left\{\left(\frac{-\alpha_1}{(1-\alpha_1^2)}\right)+(x_1-\mu_h)\frac{\alpha_0}{\sigma_v^2(1-\alpha_1)}+(x_1-\mu_h)\frac{\alpha_1x_1}{\sigma_v^2}\right\}, \\ \frac{\partial}{\partial\gamma}p_\theta(x_1) &= N_{x_1}(\mu_h,\sigma_h)\left\{(-1/(2\gamma))+(1/2)(x_1-\mu_h)^2(1-\alpha_1)^2(1/\gamma^2)\right\}\end{aligned}$$

Hence, the gradient of $p(y_2|y_1)$ can be computed. Since the gradients of $p(y_t|\mathbf{y_{t-1}})$ can be computed without changing the form of any of the requisite pdf's, the other gradient calculations follow recursively. Similarly, the Hessian of the log likelihood can also be computed. Since the optimization of the negative loglikelihood is carried out, the negative of the gradient calculation previously derived is taken. Also, the hessian calculation given below has been derived keeping the negative loglikelihood in mind. Let θ_1, θ_2 denote any of the parameters from $\theta=\{\alpha_0,\alpha_1,\sigma_v^2\}$.

$$\begin{aligned}\frac{\partial^2}{\partial\theta_1\partial\theta_2}(-\log L) &= -\frac{\partial^2}{\partial\theta_1\partial\theta_2}\sum_{t=1}^{T}\log p_\theta(y_t|\mathbf{y_{t-1}}) \\ &= -\sum_{t=1}^{T}\frac{\partial^2}{\partial\theta_1\partial\theta_2}\log p_\theta(y_t|\mathbf{y_{t-1}}) \\ &= -\sum_{t=1}^{T}\Big\{\frac{-1}{(p_\theta(y_t|\mathbf{y_{t-1}}))^{\mathbf{2}}}\left(\frac{\partial p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta_1}\right)\left(\frac{\partial p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta_2}\right)+ \\ &\quad +\frac{1}{p_\theta(y_t|\mathbf{y_{t-1}})}\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(y_t|\mathbf{y_{t-1}})\Big\} \\ &= \sum_{t=1}^{T}\Big\{\frac{1}{(p_\theta(y_t|\mathbf{y_{t-1}}))^{\mathbf{2}}}\left(\frac{\partial p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta_1}\right)\left(\frac{\partial p_\theta(y_t|\mathbf{y_{t-1}})}{\partial\theta_2}\right)-\end{aligned}$$

$$-\frac{1}{p_\theta(y_t|\mathbf{y_{t-1}})}\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(y_t|\mathbf{y_{t-1}})\Big\}$$

Again the derivation of the hessian matrix for the first time point will follow differently from the rest. The diagonal elements are given by:

$$\begin{aligned}\frac{\partial^2}{\partial\alpha_0^2}p_\theta(y_1|y_0) &= \int N_{y_1}(0,e^{x_1})\frac{\partial^2}{\partial\alpha_0^2}N_{x_1}(\mu_h,\sigma_h^2)dx_1 \\ &= \int N_{y_1}(0,e^{x_1})\left[\frac{N_{x_1}(\mu_h,\sigma_h^2)}{\sigma_h^2(1-\alpha_1)^2}\left\{\left(\frac{x_1-\mu_h}{\sigma_h}\right)^2-1\right\}\right]dx_1\end{aligned}$$

$$\begin{aligned}&\frac{\partial^2}{\partial\alpha_1^2}p_\theta(y_1|y_0) \\ =\ &\int N_{y_1}(0,e^{x_1})\frac{\partial^2}{\partial\alpha_1^2}N_{x_1}(\mu_h,\sigma_h^2) \\ =\ &\int N_{y_1}(0,e^{x_1})\frac{\partial}{\partial\alpha_1}N_{x_1}(\mu_h,\sigma_h^2)\left\{\frac{(x_1-\mu_h)(\alpha_0+\alpha_1x_1(1-\alpha_1))}{\sigma_v^2(1-\alpha_1)}-\frac{\alpha_1}{1-\alpha_1^2}\right\}dx_1 \\ =\ &\int N_{y_1}(0,e^{x_1})N_{x_1}(\mu_h,\sigma_h^2)\times \\ &\left[\frac{1}{\gamma}\left\{\frac{(x_1-\mu_h)x_1(1-2\alpha_1)}{1-\alpha_1}+(\alpha_0+\alpha_1x_1(1-\alpha_1))\left(\frac{x_1-2\mu_h}{(1-\alpha_1)^2}\right)\right\}-\frac{1+\alpha_1^2}{(1-\alpha_1^2)^2}\right]+ \\ &\left[\frac{1}{\gamma}\left\{\frac{(x_1-\mu_h)\mu_h(1-\alpha_1^2)}{1-\alpha_1}+\alpha_1(x_1-\mu_h)^2\right\}-\frac{\alpha_1}{(1-\alpha_1^2)}\right] \\ &\left\{(\alpha_0+\alpha_1x_1(1-\alpha_1))\frac{x_1-\mu_h}{\sigma_v^2(1-\alpha_1)}-\frac{\alpha_1}{1-\alpha_1^2}\right\}dx_1\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial\gamma^2}p_\theta(y_1|y_0) &= \int N_{y_1}(0,e^{x_1})\frac{\partial^2}{\partial\gamma^2}N_{x_1}(\mu_h,\sigma_h^2)dx_1 \\ &= \int N_{y_1}(0,e^{x_1})\frac{\partial}{\partial\gamma}N_{x_1}(\mu_h,\sigma_h^2)\left\{\frac{(x_1-\mu_h)^2(1-\alpha_1^2)}{2\gamma^2}-\frac{1}{2\gamma}\right\}dx_1 \\ &= \int N_{y_1}(0,e^{x_1})N_{x_1}(\mu_h,\sigma_h^2)\left[\frac{3}{4\gamma^2}-\frac{3(x_1-\mu_h)^2(1-\alpha_1^2)}{2\gamma^3}+\right. \\ &\qquad\left.+\left(\frac{(x_1-\mu_h)^2(1-\alpha_1^2)}{2\gamma^2}\right)^2\right]dx_1\end{aligned}$$

The off-diagonal elements of the hessian matrix are as follows:

$$\begin{aligned}&\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}p_\theta(y_1|y_0) \\ =\ &\frac{\partial}{\partial\alpha_1}\left[\int N_{y_1}(0,e^{x_1})N_{x_1}(\mu_h,\sigma_h^2)\frac{1}{\sigma_h(1-\alpha_1)}dx_1\right]\end{aligned}$$

$$
\begin{aligned}
= & \int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \times \\
& \left[\frac{(x_1 - \mu_h)^2}{(\sigma_h^2)^2 (1 - \alpha_1)} \left\{ \frac{(x_1 - \mu_h)\alpha_1}{(1 - \alpha_1^2)} + \frac{\mu_h}{1 - \alpha_1} \right\} - \frac{\mu_h}{\sigma_h^2 (1 - \alpha_1)^2} + \frac{(x_1 - \mu_h)(1 - 2\alpha_1)}{\sigma_h^2 (1 - \alpha_1^2)(1 - \alpha_1)} \right] dx_1
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial^2}{\partial \alpha_0 \partial \gamma} p_\theta(y_1 | y_0) \\
= & \frac{\partial}{\partial \alpha_0} \left[\int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \left\{ \frac{(x_1 - \mu_h)^2 (1 - \alpha_1^2)}{2\gamma^2} - \frac{1}{2\gamma} \right\} dx_1 \right] \\
= & \int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \frac{(x_1 - \mu_h)(1 - \alpha_1^2)}{(1 - \alpha_1)} \left[\frac{(x_1 - \mu_h)^2 (1 - \alpha_1^2)}{2\gamma^3} - \frac{3}{2\gamma^2} \right] dx_1
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial^2}{\partial \alpha_1 \partial \gamma} p_\theta(y_1 | y_0) \\
= & \frac{\partial}{\partial \gamma} \left[\int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \left\{ \frac{(x_1 - \mu_h)(\alpha_0 + \alpha_1 x_1 (1 - \alpha_1))}{\gamma (1 - \alpha_1)} - \frac{\alpha_1}{1 - \alpha_1^2} \right\} dx_1 \right] \\
= & \int N_{y_1}(0, e^{x_1}) N_{x_1}(\mu_h, \sigma_h^2) \times \\
& \left[\left\{ \frac{(x_1 - \mu_h)(\alpha_0 + \alpha_1 x_1 (1 - \alpha_1))}{\gamma (1 - \alpha_1)} - \frac{\alpha_1}{1 - \alpha_1^2} \right\} \left(\frac{(x_1 - \mu_h)^2 (1 - \alpha_1^2)}{2\gamma^2} \right) \right] \\
& - \left[\frac{1}{2\gamma} \left\{ \frac{(x_1 - \mu_h)(\alpha_0 + \alpha_1 x_1 (1 - \alpha_1))}{\gamma (1 - \alpha_1)} - \frac{\alpha_1}{1 - \alpha_1^2} \right\} \right] \\
& - \left[\frac{(x_1 - \mu_h)(\alpha_0 + \alpha_1 x_1 (1 - \alpha_1))}{\gamma^2 (1 - \alpha_1)} \right] dx_1
\end{aligned}
$$

In order to calculate the hessian matrix for the conditional probability of the second time series point, we proceed as follows:

$$
\begin{aligned}
\frac{\partial^2}{\partial \theta_1 \partial \theta_2} p_\theta(y_2 | y_1) &= \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \int p(y_2 | x_2) p_\theta(x_2 | y_1) dx_2 \\
&= \int p(y_2 | x_2) \frac{\partial^2}{\partial \theta_1 \partial \theta_2} p_\theta(x_2 | y_1) dx_2
\end{aligned}
$$

The double derivative of $p_\theta(x_2|y_1)$ is as follows:

$$
\begin{aligned}
\frac{\partial^2}{\partial \theta_1 \partial \theta_2} p_\theta(x_2 | y_1) &= \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \int p_\theta(x_2 | x_1) p_\theta(x_1 | y_1) dx_1 \\
&= \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \int \frac{p_\theta(x_2 | x_1) p(y_1 | x_1) p_\theta(x_1 | y_0) dx_1}{p_\theta(y_1 | y_0)} \\
&= \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \left(\frac{a(\theta)}{b(\theta)} \right) \\
&= \frac{\partial}{\partial \theta_2} \left\{ \frac{b(\theta) \frac{\partial}{\partial \theta_1} a(\theta) - a(\theta) \frac{\partial}{\partial \theta_1} b(\theta)}{(b(\theta))^2} \right\}
\end{aligned}
$$

$$= \frac{\partial}{\partial\theta_2}\left\{\frac{c(\theta)}{(b(\theta))^2}\right\}$$

where

$$\begin{aligned}
a(\theta) &= \int p(y_1|x_1)p_\theta(x_2|x_1)p_\theta(x_1|y_0)dx_1,\\
b(\theta) &= p_\theta(y_1|y_0),\\
c(\theta) &= b(\theta)\frac{\partial}{\partial\theta_1}a(\theta) - a(\theta)\frac{\partial}{\partial\theta_1}b(\theta)\\
&= p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_1}[\int p(y_1|x_1)p_\theta(x_2|x_1)p_\theta(x_1|y_0)dx_1]\\
&\quad -\int p(y_1|x_1)p_\theta(x_2|x_1)p_\theta(x_1|y_0)dx_1\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0),\\
\frac{c(\theta)}{(b(\theta))^2} &= \frac{1}{p_\theta(y_1|y_0)}[\int p(y_1|x_1)\frac{\partial}{\partial\theta_1}\{p_\theta(x_2|x_1)p_\theta(x_1|y_0)\}dx_1 - \{p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0)\}]\\
&= \frac{1}{p_\theta(y_1|y_0)}[\int p(y_1|x_1)[p_\theta(x_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|x_1)dx_1]\\
&\quad +\int p(y_1|x_1)[p_\theta(x_2|x_1)\frac{\partial}{\partial\theta_1}p_\theta(x_1|y_0)dx_1]\\
&\quad -\{p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0)\}]\\
&= \frac{d(\theta)}{p_\theta(y_1|y_0)}
\end{aligned}$$

Taking the second derivative of the numerator $d(\theta)$ with respect to θ_2 ie., $\frac{\partial}{\partial\theta_2}d(\theta)$:

$$\begin{aligned}
\frac{\partial}{\partial\theta_2}d(\theta) &= [\int p(y_1|x_1)\frac{\partial}{\partial\theta_2}[p_\theta(x_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|x_1)dx_1]\\
&\quad +\int p(y_1|x_1)\frac{\partial}{\partial\theta_2}[p_\theta(x_2|x_1)\frac{\partial}{\partial\theta_1}p_\theta(x_1|y_0)dx_1]\\
&\quad -\frac{\partial}{\partial\theta_2}\{p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0)\}]\\
&= \int p(y_1|x_1)\{\frac{\partial}{\partial\theta_2}p_\theta(x_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|x_1) + p_\theta(x_1|y_0)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_2|x_1)\}dx_1\\
&\quad +\int p(y_1|x_1)\{\frac{\partial}{\partial\theta_2}p_\theta(x_2|x_1)\frac{\partial}{\partial\theta_1}p_\theta(x_1|y_0) + p_\theta(x_2|x_1)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_1|y_0)\}dx_1\\
&\quad -\{\frac{\partial}{\partial\theta_2}p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0) + p_\theta(x_2|y_1)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(y_1|y_0)\}
\end{aligned}$$

Hence, after taking the second derivative:

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_2|y_1) = \frac{\partial}{\partial\theta_2}\left\{\frac{d(\theta)}{p_\theta(y_1|y_0)}\right\} = \frac{\frac{\partial}{\partial\theta_2}d(\theta)}{p_\theta(y_1|y_0)} - \frac{d(\theta)\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2}$$

However, the second term in the above expression can be simplified further:

$$\begin{aligned}
\frac{d(\theta)\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2} &= \frac{\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2}\int p(y_1|x_1)\frac{\partial}{\partial\theta_1}[p_\theta(x_1|y_0)p_\theta(x_2|x_1)]dx_1 \\
&\quad -\frac{p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2} \\
&= \frac{\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2}\frac{\partial}{\partial\theta_1}[p_\theta(x_2|y_1)p_\theta(y_1|y_0)] \\
&\quad -\frac{p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)}{(p_\theta(y_1|y_0))^2} \\
&= \frac{\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|y_1)}{p_\theta(y_1|y_0)}
\end{aligned}$$

Hence, substituting this in the above equation:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_2|y_1) &= \frac{1}{p_\theta(y_1|y_0)}[\frac{\partial}{\partial\theta_2}d(\theta) - \frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|y_1)] \\
&= \frac{1}{p_\theta(y_1|y_0)}\Big[\int p(y_1|x_1)\{\frac{\partial}{\partial\theta_2}p_\theta(x_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|x_1) \\
&\quad +p_\theta(x_1|y_0)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_2|x_1)\}dx_1 \\
&\quad +\int p(y_1|x_1)\Big[\frac{\partial}{\partial\theta_2}p_\theta(x_2|x_1)\frac{\partial}{\partial\theta_1}p_\theta(x_1|y_0) + \\
&\quad p_\theta(x_2|x_1)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_1|y_0)\Big]dx_1 \\
&\quad -\{\frac{\partial}{\partial\theta_2}p_\theta(x_2|y_1)\frac{\partial}{\partial\theta_1}p_\theta(y_1|y_0) + p_\theta(x_2|y_1)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(y_1|y_0)\} \\
&\quad -\frac{\partial}{\partial\theta_2}p_\theta(y_1|y_0)\frac{\partial}{\partial\theta_1}p_\theta(x_2|y_1)\Big]
\end{aligned}$$

In order to compute the above, the calculation of the double derivative of $p(x_2|x_1)$ remains:

$$\begin{aligned}
\frac{\partial^2}{\partial\alpha_0^2}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\left\{\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{(\sigma_v^2)^2} - \frac{1}{\sigma_v^2}\right\} \\
\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\left\{\frac{x_1(x_2-\alpha_0-\alpha_1x_1)^2}{(\sigma_v^2)^2} - \frac{x_1}{\sigma_v^2}\right\} \\
\frac{\partial^2}{\partial\alpha_0\partial\gamma}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\left\{\frac{(x_2-\alpha_0-\alpha_1x_1)^3}{2(\sigma_v^2)^3} - \frac{3(x_2-\alpha_0-\alpha_1x_1)}{2(\sigma_v^2)^2}\right\} \\
\frac{\partial^2}{\partial\alpha_1^2}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\left\{\frac{x_1^2(x_2-\alpha_0-\alpha_1x_1)^2}{(\sigma_v^2)^2} - \frac{x_1^2}{\sigma_v^2}\right\} \\
\frac{\partial^2}{\partial\alpha_1\partial\gamma}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\left\{\frac{x_1(x_2-\alpha_0-\alpha_1x_1)^3}{2(\sigma_v^2)^3} - \frac{3x_1(x_2-\alpha_0-\alpha_1x_1)}{2(\sigma_v^2)^2}\right\}
\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial\gamma^2}p_\theta(x_2|x_1) &= N_{x_2}(\alpha_0+\alpha_1x_1,\sigma_v^2)\times \\ &\times\left[\left(\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{2(\sigma_v^2)^2}-\frac{1}{2\sigma_v^2}\right)\frac{(x_2-\alpha_0-\alpha_1x_1)^2}{2(\sigma_v^2)^2}-\right. \\ &\left.-\frac{5(x_2-\alpha_0-\alpha_1x_1)^2}{4(\sigma_v^2)^3}+\frac{3}{4(\sigma_v^2)^2}\right]\end{aligned}$$

Hence, the final step in the calculation of the hessian of $p(y_2|y_1)$ is:

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(y_2|y_1) = \int p(y_2|x_2)\frac{\partial^2}{\partial\theta_1\partial\theta_2}p_\theta(x_2|y_1)dx_2$$

The previous recursive calculation stays unaltered for calculating the gradient and the hessian of $p(y_t|\mathbf{y_{t-1}})$ for $t=2,3,\ldots T$.