7-26-2021 10:00 AM

On the Estimation of Heston-Nandi GARCH Using Returns and/or Options: A Simulation-based Approach

Xize Ye, *The University of Western Ontario*

Supervisor: Escobar-Anel, Marcos, *The University of Western Ontario*
Co-Supervisor: Stentoft, Lars, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

Abstract

In this thesis, the Heston-Nandi GARCH(1,1) (henceforth, HN-GARCH) option pricing model is fitted via 4 maximum likelihood-based estimation and calibration approaches using simulated returns and/or options. The purpose is to examine the benefits of the joint estimation using both returns and options over the fundamental returns-only estimation on GARCH models. From our empirical studies, with the additional option sample, we can improve the efficiency of the estimates for HN-GARCH parameters. Nonetheless, the improvements for the risk premium factor, both from empirical standard errors, and sample RMSEs, are insignificant. In addition, option prices are simulated with a pre-defined noise structure and with different noise levels, to demonstrate the consequence when we have a noisy option sample versus a less noisy one. The result shows that, with added option samples the RMSEs for estimated GARCH parameters are reduced dramatically, even with a very noisy option data set. This suggests that calibrating GARCH option pricing models with a relatively short return series of around 6 years, plus an option sample is more ideal than using a long return series of 20 years alone. Finally, as a by-product, we studied which type of options leads to the larger calibration improvements. Our controlled experiment confirms that out-of-the-money, short-maturity options are the best choices.

Keywords: Heston-Nandi GARCH, Simulation, Joint Calibration-Estimation, Maximum Likelihood Estimation

Summary for Lay Audience

The development of option pricing models has been a productive research area ever since the first Nobel prize-winning proposal of the Black-Scholes-Merton model in the 1970s. In 2000, Heston and Nandi proposed a particular GARCH(p, q) conditional volatility model with a closed-form option pricing formula for European option prices. The filtering and estimation of conditional volatility of this model can be completed solely from daily observables. Yet, many questions remain unanswered. For instance, is the model calibration process robust and reliable? How accurate and valid are the parameter estimates? In this thesis, we simulated market data based on the Heston Nandi-GARCH(1,1) model, and then examined and compared the 4 maximum likelihood-based estimation and calibration approaches using returns and/or options. We first followed Bollerslev (1986) and Heston and Nandi (2000) to investigate the fundamental returns-only estimation on GARCH models, during which we found that the price of risk parameter, is particularly difficult to estimate from returns data only, and its estimator is highly influenced by the average level of simulated daily noises. Hence, we simulated option prices with a pre-defined noise structure, calibrated the model jointly with options data, and compared its performance with the benchmark returns-only MLE method. We conjectured beforehand that bringing in option data shall help calibrate all parameters, with the potential of capturing the risk premium more precisely. From our empirical studies, with the additional option sample, we can improve the efficiency of the estimates for HN-GARCH parameters. Nonetheless, the improvements for the risk premium factor, both from empirical standard errors, and sample RMSEs, are insignificant. In addition, we simulated the option sample with different noise levels, to demonstrate the consequence when we have a noisy option sample versus a less noisy one. The result shows that, with added option samples the RMSEs for estimated GARCH parameters are reduced dramatically, even with a very noisy option data set. This suggests that calibrating GARCH option pricing models with a relatively short return series of around 6 years, plus an option sample is more ideal than using a long return series of 20 years alone. Finally, as a by-product, we studied which type of options leads to the larger calibration improvements. Our controlled experiment confirms that out-of-the-money, short-maturity options are the best choices.

Acknowledgements

I would like to thank my supervisors, Professor Marcos Escobar-Anel, and Professor Lars Stentoft, for their supervision and support. Professor Escobar is the chair of the financial modelling graduate program at Western University, who originally recommended my admission to the graduate program, and hence provided me a valuable opportunity to continue my academic path. During the time working as a research assistant, Professor Escobar and Professor Stentoft provided creative advice and suggestions, and also encouraged me when I felt lost. Although being extremely busy, they still found time to provide detailed feedback and helpful comments on my work. I felt touched by the kind, delightful personalities that both professors share, and I was motivated by their conscientious attitudes towards academic research. I am very grateful for the assistance from both professors, without which this work will not be present.

I also want to thank my parents, Xiaojuan Ding and Guoxing Xu, for their continuous love, support, and understanding. Their consistent support has been, and will always be my company all along the journey.

Dedication

This work is dedicated to my grandfather Bingshan Ding and my grandmother Shuzhen Lv, who raised me when I was little. Both of you have been amazing mentors that not only took great care of me throughout my childhood, but also gave me unconditional love and support. The completion of the thesis is a milestone in my academic career. Thank you for everything, and I wish you were there.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, we have witnessed great developments of option pricing models that incorporate different stylized facts observed in financial markets. Researchers tend to spot the disparity between model assumptions and realities, and build upon previous models. As an example, Heston proposed a continuous-time stochastic volatility model that replaced the constant volatility term σ in the fundamental Black-Scholes-Merton (henceforth B-S) model by the famous Cox-Ingersoll-Ross (CIR) process (see Heston, 1993). The resulting stochastic volatility model can capture the correlation between spot asset return with conditional volatility, while remaining an affine structure in order to efficiently price options. However, continuous-time stochastic volatility models often contain latent conditional volatility dynamics, hence it is difficult to estimate volatility parameters using discretized asset prices. To overcome this issue, Heston and Nandi (2000) developed a closed-form option pricing model where the spot asset price follows a particular GARCH(p, q) process. Thanks to its configuration, the HN-GARCH model captures stylized facts of seasonal conditional volatility, yet much easier to implement and test, due to its discrete-time feature with a 1-day lag.

The flexibility of HN-GARCH motivated several extensions. We have seen that Christoffersen, Heston, and Jacobs (2013) equipped the original HN-GARCH model with a variance-dependent pricing kernel by adding another parameter to explain some stylized market anomalies (such as the U-shaped pricing kernel) hence allowing better empirical fitting. Moreover, Orthanalai (2014) proposed a GARCH process by adding Levy jumps to the HN-GARCH model to capture the non-Gaussianity of asset returns, and argued that the infinite-activity jumps are more representative of shocks in asset prices than the Brownian increment. Very recently, Escobar, Rastegari, and Stentoft (2019) proposed a multivariate analogy of the HN-GARCH model to allow correlation between multiple assets in pricing multi-asset options, while maintaining the affine structure. Undoubtedly, these extensions open up new horizons for GARCH option valuation models and yielded rich applications. Nonetheless, the literature in this field mostly focuses on addressing the shortcomings of current models and exploring their fixes, without giving attention to the equally important question regarding the validity and accuracy of the model estimation and calibration process.[1] In particular, the literature regarding the model always includes the model fitting results with real market data, from which one does not know the true parameters, hence creating uncertainty on the validity of the finite sample estimates. Therefore, in this thesis, we hope to fill the gap to investigate the performance and properties of the popular calibration and estimation approaches, with the help of simulated market data. We focus on the 4 popular approaches to estimate the model parameters for option pricing models:

1. Maximum likelihood estimation with stock returns data only (see Bollerslev, 1986 on estimating GARCH models, and Heston and Nandi, 2000 on estimating HN-GARCH). This is the most fundamental method to fit GARCH models, and so this approach is called "Returns-only (MLE) estimation".

2. Non-linear least square approach that fits the model implied option prices with the observed prices via the vega-weighted Gaussian likelihood function, where the daily conditional variances are kept fixed. We name this approach "Options-only calibration method 1".

3. Non-linear least square approach that fits the model implied option prices with the observed prices via the vega-weighted Gaussian likelihood function, where the daily conditional variances are updated from observed returns. This method is comparable to Options-only calibration method 1, however, we would also need return observations. In each iteration of the numerical optimization process, we compute the implied conditional variances, based on which we compute model option prices. This particular method of filtering the variances from returns is popular in existing literature, including Heston and Nandi (2000), Christoffersen and Jacobs (2004), and many others. Combining further with the vega-weighted Gaussian likelihood function, we name this approach

[1] Throughout the thesis, we use "estimation" to refer to estimating the model parameters via asset return observations, and we use "calibration" to refer to estimating model parameters via option prices.

"Options-only calibration method 2".[2]

4. Joint maximum likelihood estimation that estimates the parameters by maximizing the combined return and option likelihood, whereas in "Options-only calibration method 2", we also filter conditional variances from returns. This method is largely seen in recent works (for example, in Kanniainen, Lin, and Yang, 2014). The only difference between the joint estimation and the "Options-only calibration method 2" approach is that the latter only maximizes the option likelihood alone, where the joint approach contains the return likelihood. We call this approach "Joint returns-options estimation-calibration".

The idea of using market option prices in model calibration, in both using options alone, or jointly with return data to further enhance the accuracy of parameter estimates has been widely applied. However, the existing literature is particularly silent about the properties and performance of such joint estimation, when the data is generated by the true model. In particular, what are the finite sample and asymptotic properties of the estimation procedure? Do we have a conclusion when we compare the various estimation approaches using returns and options? Such questions cannot be answered without simulation studies, which motivates our work.

The main objective of this thesis is twofold. First, we examine the validity and difficulty of the model estimation and calibration process with the help of simulated market return and option data. In this thesis, in addition to the assumed return structure of HN-GARCH, we fix a statistical structure for observational noises of option prices, which leads to inference methodologies utilizing return and option observations. We first present a detailed explanation of the financial intuition behind the model and the parameters, and we give a brief introduction to model calibration methodologies, for instance, the MLE procedure and the standard error estimates for parameters. Since the end goal is to study the properties and performance of the joint estimation using both returns and options, we conducted extensive simulation-based studies by first generating 100 samples for various sample sizes of stock returns and option prices, optimizing each sample, and then extracting basic statistics, such as RMSEs, and empirical distributions of the parameter estimates. We focus on

[2] In the literature, this approach is rarely used alone, as the standard approach (as seen in Christoffersen, Heston, and Jacobs, 2013, Kanniainen, Lin, and Yang, 2014, and many others) will combine the option likelihood with the return likelihood for a joint estimation (see point 4). From practical perspectives, such joint estimation approach should always be considered over the options-only approach. Nonetheless, since part of our objective is to understand the information and predictability of the model parameters contained in the option prices, we mention and use this approach in our studies, and further compare its estimating result with the joint estimation result.

confirming convergence, and finite-sample properties of the maximum likelihood estimator and its standard error estimates, and concluded that the above 4 estimation and calibration procedures are indeed valid. Throughout the process, we also discuss some of the numerical challenges and their potential fixes.

Second, we compare the above estimation and calibration methods to examine the benefit of the Options-only calibration and Joint estimation-calibration approaches over the benchmark Returns-only estimation approach. The HN-GARCH model contains 5 parameters, where λ is the risk premium factor, and ω, α, β, and γ are GARCH parameters that governs the conditional variances (see Section 2). Based on our previous findings, the risk premium parameter, λ, cannot be effectively estimated from return data only, since the standard error for λ and the estimate of λ are of similar sizes in Returns-only MLE with $N = 4500$ (see Table 4.2 for example). Hence, although we compare the estimation result for all the parameters, we are curious whether the calibration methods involving options have any benefits on λ. To carry out the comparison, we compare 12 groups of Joint estimation-calibration with different numbers of returns and options in the sample (hence different percentage of options, from 0% to 90%, where 0% of options refers to the Returns-only estimation) to validate the importance of options in the joint estimation.

As a summary, we outline the main contributions of our work:

1. Inspired by practical implementations in the literature, we assume observed option prices follow the true model prices plus vega weighted normal noises. The details of this particular structure is included in Chapter 2. This setup allows us to do a joint maximum likelihood-based inference on Heston-Nandi stock returns and option prices, with details explained in Chapter 3.

2. In Chapter 4, we demonstrated the validity of such joint estimation via a simulation study, where the model is the true data generating process. We demonstrated consistency of both the parameter estimators, and the standard error estimates based on the outer product of gradients.

3. Also in Chapter 4, the advantages of the Joint estimation-calibration are presented in comparison to Options-only calibration and to Returns-only MLE. In particular, we demonstrate the impact of increasing sample sizes (returns and options), and with different variances of options' vega errors (i.e. when the option sample is more noisy and less noisy). The result shows that the gains in RMSEs and standard errors for GARCH parameters α, β, and γ when we increase either the return or option observations is noticeable. However, the

change in RMSEs for λ is only significant when we increase the sample size of returns.

4. As a by-product, we studied which type of options leads to the larger calibration and estimation-calibration improvements by comparing the sample RMSEs of option samples consisting of different types of options. Our controlled experiment confirms that out-of-the-money, short maturity call options are the best choices. The results and discussions are in Appendix A.

The structure of the remaining thesis is as follows: Chapter 2 and Chapter 3 describe basic properties of the model and statistical inference methods. Chapter 4 describes the data generating process, and discusses the numerical estimation and calibration results. Chapter 5 discusses current problems, concludes, and gives directions for future work. The appendix includes work in progress, and also supporting information and results. In particular, Appendix A discusses which type of options, in terms of moneyness and maturity, is the best in the calibration with options only. In Appendix B, we present the results from the same experiment, but with a different set of true parameters, to confirm the main results.

Chapter 2

HN-GARCH Model Description

We begin with a detailed introduction to the HN-GARCH model. In this chapter, we state the dynamics of the return and conditional variance processes under physical and risk-neutral measures. We further outline the model assumptions and some properties, including the closed-form option pricing formula, and explain the intuitions and restrictions behind each model parameter.

2.1 Assumptions

We begin by stating the two assumptions used in HN-GARCH. The first assumption directly gives the historical process of the risky asset in the market while the second assumption implies the risk-neutral process.

Assumption 2.1. The Heston-Nandi GARCH(1,1) model assumes the spot asset price follows a GARCH-type process specified as

$$\log(S(t)) = \log(S(t-1)) + r + \lambda h(t) + \sqrt{h(t)}z(t) \tag{2.1}$$

$$h(t) = \omega + \beta h(t-1) + \alpha \left(z(t-1) - \gamma\sqrt{h(t-1)} \right)^2, \tag{2.2}$$

where $t = 0, 1, ...,$ S_t is the stock price at time t, r is the continuously compounded interest rate, $h(t)$ (known at $t-1$), is the variance of $S(t)$ given all the information up to $t-1$, and $z(t)$ is a sequence of independent and identically distributed (i.i.d.) standard normal noises. We will impose conditions later to ensure the conditional variance process $h(t)$ is stationary with finite mean.

Besides the setting for the physical distribution, one also needs assumptions for the stock process under the risk-neutral measure. A common way to deduce the risk-neutral dynamics is by stating a (any) pricing kernel and use the implied risk-neutral process (see Christoffersen, Heston, and Jacobs (2013) for an example of using a variance-dependent pricing kernel). In this thesis, since the focus is to study the theoretical benefits of bringing in options into calibration, we will follow Heston and Nandi (2000)'s original assumption for simplicity.

Assumption 2.2. The single period European call option price with strike price K and spot asset price S_t obeys the Black-Scholes formula

$$C(S_t, K, T, \sigma) = S_t N(d_1) - Ke^{-rT} N(d_1) - \sigma\sqrt{T} \tag{2.3}$$

$$d_1 = \frac{\ln(S_T/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{2.4}$$

where $T = 1/252$, σ is the annualized spot conditional volatility, and $N(\cdot)$ is the cumulative density function for a standard normal variable.

Assumption 2.2 implies that the risk-neutral expectation of $S(t+1)$ given $S(t)$ equals $S(t)e^r$, hence, by merely applying some transformations to replace λ by $-1/2$ and γ by $\gamma^* = \gamma + \lambda + 1/2$, we realize the risk-neutral process takes the same GARCH form as

$$\log(S(t)) = \log(S(t-1)) + r - \frac{1}{2}h(t) + \sqrt{h(t)}z^*(t) \tag{2.5}$$

$$h(t) = \omega + \beta h(t-1) + \alpha\left(z^*(t-1) - \gamma^*\sqrt{h(t-1)}\right)^2, \tag{2.6}$$

where $z^*(t)$ is the i.i.d. standard normal noises under the risk-neutral measure (see Heston and Nandi (2000) for proof and explanations of this implication).

By the previous assumption, we can price European options with any maturity. However, in our calibration process, we also need to give distributional assumptions to the observational errors, in order to draw inference such as deriving standard errors for our parameter estimation. Many researchers tend to give financial interpretations of the existence of such observational errors on market option prices, for example, Hentschel (2003) discussed 3 major sources of observational errors: errors due to finite quote precisions; errors from bid-ask spread, and errors from the misuses of non-synchronous prices and their magnitudes. Also, Jacquier and Jarrow (2000) claimed that market option price contains errors from mispricing or inaccurate observations, and gave a statistical framework that the observed prices equal a log-normally distributed error coefficient multiplied by the true price. We will rely on the widely popular vega weighted MSE metric to assume that the vega weighted option errors follow a normal random variable with mean 0, as stated in the next assumption.

Assumption 2.3. The vega weighted option errors

$$e_i = (C_i - C_i(\boldsymbol{\theta}))/\nu_i, \tag{2.7}$$

are independent and identically distributed normal random variables (which are also independent of the daily return innovations z_i in (2.1)) with mean 0 and variance σ^2, where C_i is the ith observed option price, $C_i(\theta)$ is the ith true option price given by the model (see Section 2.3 for its representation) with a vector of parameters $\boldsymbol{\theta}$, and ν_i is the Black-Scholes vega computed at the market level of implied volatility.

As standard in the finance literature, the popular use of the vega weighted Gaussian log-likelihood function is merely from the assumption that the Black-Scholes implied volatility of observed option prices follows the implied volatility of true option prices plus normal noises. Because inverting the Black-Scholes formula is computationally expensive, we use the linear vega estimate for the difference in implied volatility. Examples of the prevalent uses of such objective function are seen in Christoffersen and Jacobs (2001), the Bank of Canada working paper by Christoffersen, Feunou, Jacobs, and Meddahi (2012), and Kanniainen, and Lin and Yang (2014). With the presence of this assumption, the market observables, namely, historical asset prices and option prices, are solely based on model parameters, plus structured random noises. Furthermore, the (Gaussian) likelihood of single return or option observation can be computed effortlessly. Therefore, with the help of Assumption 2.3, the simulated option prices is consistent with the standard calibration approaches in the literature.

2.2 Properties of HN-GARCH

We outline some of the basic properties implied by equations (2.1) and (2.2). Note that because the physical and risk-neutral process takes the same GARCH form, and the only differences are the parameter replacement on λ and γ, we only outline the properties based on physical processes. We can always alter λ and γ to get the properties for the risk-neutral processes. We begin by computing some basic statistics.

Proposition 2.4 (Conditional distribution of $S(t+1)$)**.** *Conditional on all the information up to $\mathcal{F}_t$, the next-day asset price $S(t+1)$ is a log-normal random variable specified as:*

$$S(t+1)|\mathcal{F}_t \sim LN\left(\log(S(t)) + r + \lambda h(t+1), h(t)\right). \tag{2.8}$$

The proof is trivial by noting that $h(t+1)$ is known at t. Hence, we see that $\log(S(t+1))|\mathcal{F}_t \sim N(\log(S(t-1)) + r + \lambda h(t), h(t))$. This proposition has some interesting implications:

1. The next-day stock price conditioning on the previous day is log-normally distributed. This is consistent with the Black-Scholes setup.

2. In contrast to traditional Black-Scholes, the model uses a specific GARCH process for conditional volatility. As a result, we would have a (stochastic) discrete-time series $h(t)$ specifying the daily conditional variance. From (2.8), we see that $h(t)$ is indeed the daily conditional variance.

3. The log stock price $\log S(t+1)$ conditional on $\mathcal{F}_t$ has an expected value of $\log S(t)+r+\lambda h(t)$. Hence, based on our specification, the degree of conditional volatility has an impact on the average 1-day stock return through the risk premium parameter λ. When $\lambda > 0$, then per unit risk the expected next-day log stock price would increase by λ. In reality, we would often see that a more volatile stock has a higher price. This is because higher volatility implies a higher probability for the stock price to rise, hence resulting in higher extreme gains. Although on the other hand, the probability for extreme losses is also bigger due to higher volatility, the maximum loss is bounded by the current stock price $S(t)$, hence on average, we should expect stocks with higher volatility to have higher returns, which suggests that $\lambda > 0$. On the other hand, for risk-less assets, the terms $\lambda h(t)$ and $\sqrt{h(t)}z(t)$ will vanish hence the next-day asset price $S(t+1)$ will just be today's price $S(t)$ multiplied by e^r. Thus the introduction of λ does not induce arbitrage problems.

   As addressed by Duan (1995), although many empirical observations show evidence of the appearance and importance of the risk premium factor (λ in the HN-GARCH context), most models do not have this configuration. Hence, we emphasize the importance of λ and its implications in option pricing here.

We now introduce two propositions that focus on the conditional volatility process $h(t)$.

Proposition 2.5 (Expectation of $h(t)$)**.** *With the standard GARCH assumptions on $h(t)$, we have the following:*

a) The long-run unconditional daily variance, $\mathbb{E}(h(t))$, equals

$$\mathbb{E}(h(t)) = \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2}, \tag{2.9}$$

from which we can approximate the expected annual volatility of stock

$$\sigma_{\text{annual}} = \sqrt{252 \times \mathbb{E}(h(t))} = \sqrt{252 \times \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2}}. \tag{2.10}$$

b) The conditional next-day variance is a linear function of today's variance:

$$\mathbb{E}_{t-1}(h(t+1)) = (\beta + \alpha\gamma^2)h(t) + (1 - \beta - \alpha\gamma^2)\mathbb{E}(h(t)). \tag{2.11}$$

As a consequence, the conditional variance $h(t)$ reverts to its unconditional level exponentially, with a rate of $1 - \beta - \alpha\gamma^2$.

c) The conditional variance $h(t)$ and asset log return are correlated as

$$\text{Cov}_{t-1}\left(h(t+1), \log(S(t)/S(t-1))\right) = -2\alpha\gamma h(t). \tag{2.12}$$

Proof. We first expand the brackets in the expression of $h(t)$ to get

$$\begin{aligned} h(t+1) &= \omega + \beta h(t) + \alpha\left(z(t) - \gamma\sqrt{h(t)}\right)^2 \\ &= \omega + \beta h(t) + \alpha\left(z^2(t) + \gamma^2 h(t) - 2\gamma\sqrt{h(t)}z(t)\right) \\ &= \omega + \beta h(t) + \alpha z^2(t) + \alpha\gamma^2 h(t) - 2\alpha\gamma\sqrt{h(t)}z(t). \end{aligned}$$

Taking expectation (conditional on $\mathcal{F}_{t-1}$) on both sides gives (note that $h(t)$ is known at $t-1$)

$$\begin{aligned} \mathbb{E}_{t-1}(h(t+1)) &= \omega + \beta h(t) + \alpha\mathbb{E}_{t-1}(z^2(t)) + \alpha\gamma^2 h(t) - 2\alpha\gamma\sqrt{h(t)}\mathbb{E}_{t-1}(z(t)) \\ &= \omega + \beta h(t) + \alpha + \alpha\gamma^2 h(t). \end{aligned}$$

Taking expectation on both sides again (and by tower property) gives

$$\begin{aligned} \mathbb{E}(h(t)) = \mathbb{E}(h(t+1)) &= \mathbb{E}(\mathbb{E}_{t-1}(h(t+1))) \\ &= \omega + \alpha + (\beta + \alpha\gamma^2)\mathbb{E}(h(t)), \end{aligned}$$

so

$$\mathbb{E}(h(t)) = \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2},$$

as seen in a). Also, re-arranging terms in the second equality gives

$$\begin{aligned}\mathbb{E}_{t-1}(h(t+1)) &= (\beta+\alpha\gamma^2)h(t)+\omega+\alpha\\ &= (\beta+\alpha\gamma^2)h(t)+(1-\beta-\alpha\gamma^2)\mathbb{E}(h(t)),\end{aligned}$$

which shows b). To prove c), one have to note that

$$\mathrm{Cov}_{t-1}(X,Y)=0$$

whenever X is $\mathcal{F}_{t-1}$-measurable. Hence, by bi-linearity of conditional covariance, we have

$$\begin{aligned}\mathrm{Cov}_{t-1}\left(h(t+1),\log\left(\frac{S(t)}{S(t-1)}\right)\right) &= \mathrm{Cov}_{t-1}\left(h(t+1), r+\lambda h(t)+\sqrt{h(t)}z(t)\right)\\ &= \sqrt{h(t)}\,\mathrm{Cov}_{t-1}\left(h(t+1), z(t)\right)\\ &= \sqrt{h(t)}\,\mathrm{Cov}_{t-1}\left(\alpha\left(z^2(t)-2\gamma\sqrt{h(t)}z(t)\right), z(t)\right),\end{aligned}$$

where we note that

$$\mathrm{Cov}_{t-1}\left(z^2(t), z(t)\right)=\mathbb{E}(z^3(t))-\mathbb{E}(z^2(t))\mathbb{E}(z(t))=0,$$

thus,

$$\begin{aligned}\mathrm{Cov}_{t-1}\left(h(t+1),\log(S(t)/S(t-1))\right) &= \sqrt{h(t)}\,\mathrm{Cov}_{t-1}\left(\alpha-2\gamma\sqrt{h(t)}z(t), z(t)\right)\\ &= -2\alpha\gamma h(t)\,\mathrm{Cov}(z(t),z(t))\\ &= -2\alpha\gamma h(t).\end{aligned}$$

□

These properties also have meaningful implications:

Property a) helps us calculate the long-run daily variance, as well as the annualized variance and volatility given a set of parameters. This is particularly helpful in the data generation process, as they provide basic statistics of the generated return series.

Property b) shows the exponential mean-reverting property of conditional variance process $h(t)$. In model calibration using return data, assume we observe a vector of stock prices. In order to compute the log-likelihood vector given a set of parameters, we need the initial volatility $h(1)$ to find the implied daily noises recursively. Because of the mean-reverting property, the backward computation of the vector $z(t)$ and $h(t)$ is not heavily dependent on the choice of initial variance $h(1)$. In other words, depending on the level of persistence $(\beta + \alpha\gamma^2)$, after hundreds of iterations, the conditional variances $h(t)$ will be close to the true variances once we are given the true set of parameters. Hence, as long as we drop a leading portion of the data, the calibration result will be insensitive to the choice of initial volatility $h(1)$. A more detailed discussion is presented in Chapter 4.

Property c) tells us the configuration of HN-GARCH allows arbitrary correlation between asset return and volatility. In particular, if $\alpha = 0$, then the volatility is a non-random, time-varying process. Hence, there is no correlation between volatility and return. However, if $\alpha > 0$, then a positive γ brings a negative correlation between asset log return and volatility, which is consistent with empirical findings captured by Black (1976) and Christie (1982), commonly known as the leverage effect. Also, Campbell and Hentschel (1992) and Bekaert and Wu (2000) find evidence of the volatility feedback effect (or news effect) that also confirms such negative correlation (see Bae, Kim, and Nelson (2004) for a discussion).

The GARCH process given by equations (2.1) and (2.2) assumes discrete observations of asset prices with an interval of 1 unit of time. If we replace the prescribed interval by an arbitrary amount, denoted by Δ, then we have the following result when Δ shrinks to 0. As $h(t)$ is the conditional variance of returns for the next Δ period, when Δ shrinks we must see $h(t)$ converging to 0. Hence we define the variance per unit time $v(t) = h(t)/\Delta$ and we can formulate limiting properties of $v(t)$. Because the variance and return process ($h(t)$ and $R(t)$) of HN-GARCH depends on the same noise $z(t)$, its continuous-time limit remains the same structure, where the dynamics $d \log S(t)$ and $dv(t)$ depends on the same Wiener process $W(t)$, resulting in a special Heston model where the log return and variance processes are perfectly correlated.

Proposition 2.6 (Continuous-time limit)**.** *If the time interval Δ shrinks to 0, then the stochastic process $v(t)$ converges to a special Cox-Ingersoll-Ross (CIR) process, that is driven by the same Brownian motion as the continuous-time limit of log asset prices. As a consequence, the HN-GARCH model contains the perfectly correlated Heston (1993) stochastic volatility model as the continuous-time limit.*

Proof. See Heston and Nandi (2000) Appendix B. Convergence to Continuous Time. □

2.2.1 Parameter intuition

In the HN-GARCH model, the parameter vector $\boldsymbol{\theta}$ contains 5 distinct parameters, namely,

$$\boldsymbol{\theta} = (\lambda, \omega, \alpha, \beta, \gamma). \tag{2.13}$$

We have seen earlier that each parameter has its individual interpretation and intuition. In this section, we focus on describing the implication and meaning of each parameter.

- λ: Parameter λ is also referred to as the risk premium. We have discussed earlier that the asset with higher volatility usually has a higher price, due to higher expected extreme gain. The parameter λ governs, how much per unit variance, the physical average of log asset return increases. The fact that risky assets have higher average returns than risk-free assets is not only from economical reasoning, but also confirmed by empirical evidence. By having the risk premium parameter in HN-GARCH, the model fits better, especially with long return series.

- ω and β: The volatility process $h(t)$ has an autoregressive component, where part of the next-day variance, $h(t+1)$, equals a fixed number (ω) plus a percentage of previous day's variance ($\beta h(t)$). It is intuitively clear that the conditional variances of two consecutive days are highly positively correlated. If the market is highly volatile today, then with a very high probability the next day will remain the same. The parameter β, usually close to 1, suggests that a considerable portion of today's variance goes to the next day, while ω is constantly added on each day.

- α and γ: The parameter α determines the heaviness of the tail distribution of $h(t)$. Because α is the coefficient of the random portion of $h(t)$, a bigger coefficient gives a heavier tail and vice versa. In the extreme case where $\alpha = 0$, then the process $h(t)$ is non-random, hence no tail. On the other hand, γ controls the level of asymmetry of the distribution of $h(t)$. If $\gamma = 0$, then the resulting distribution is symmetrical, because the extreme positive or negative values of daily noise have equal effects on $h(t)$. However, when $\gamma > 0$, then a large negative noise is more influential than a large positive noise with an equal amount, hence creating asymmetry. Also, parameters α and γ govern jointly the correlation between asset log return and volatility, as seen in (2.12). Hence, with a positive α, $\gamma > 0$ implies a negative correlation, which is consistent with empirical findings.

Besides the effects introduced above, it should be mentioned that these parameters determine all the moments of the conditional variance $h(t)$. Among these useful statistics, we only include the derivation of persistence and mean in Proposition 2.5. The formulas for the first four moments can be found in the appendix in Lahouel and Hellara (2017). Next, we show how each parameter influences the physical stock price series by comparing two parameter sets.

2.2.2 A comparison of parameter set

Different choices of parameters can have a huge impact on the model, since the parameters fully characterize the stock returns and determine the moments and correlations. In our work, we want the simulated data to be close to reality, we therefore choose three sets of parameters that are the maximum likelihood estimation results in Heston and Nandi (2000), Christoffersen, Heston and Jacobs (2006), and Christoffersen, Heston and Jacobs (2013). Henceforth, we will refer to these sets of parameters as the H-N 2000, C-H-J 2006, and C-H-J 2013 parameters. We now examine the effect of these parameter sets.

Heston-Nandi 2000 parameters From Heston and Nandi (2000), we obtain the MLE estimates for the parameters using daily S&P 500 cash index level from January 8, 1992 to December 30, 1994 for a total of 755 observations.[1] The estimated parameters are:

$$\begin{cases} \lambda = 0.205, \\ \omega = 5.02 \times 10^{-6}, \\ \alpha = 1.32 \times 10^{-6}, \\ \beta = 0.589, \\ \gamma = 421.39, \end{cases} \tag{2.14}$$

with these parameters, and

$$r = 0, \;\; S(0) = 100, \;\; h(1) = 3.59e\text{-}5, \tag{2.15}$$

we are able to generate (recursively) arbitrarily long sample paths.[2] The following graph shows 4 different sample trajectories generated by the parameters in (2.14) and (2.15) with $N = 5000$.

[1]The daily cash level closest to (before) 2:30 P.M. (central standard time) on each day are used.

[2]The value $3.59e$-5 is the long-run average conditional variance for parameters from (2.14). The daily disturbances are generated from Matlab's random number generator for i.i.d. standard normal distributions.

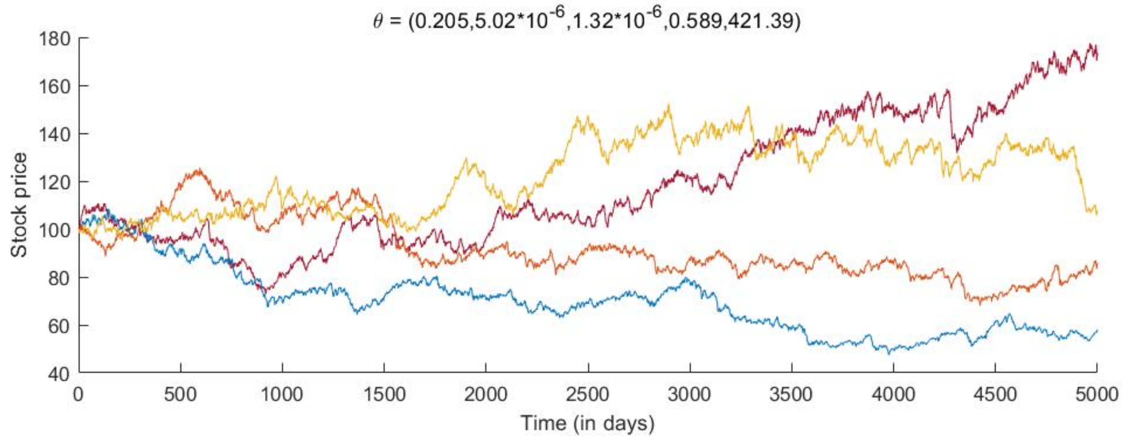


Figure 2.1: Sample trajectories for $N = 5000$, with parameters in (2.14) and (2.15)

With this set of parameters, the estimated annual expected return equals 0.634%, and the annual expected volatility is 9.491%, both calculated via Monte Carlo simulation with $M = 100,000$ trajectories. The theoretical annual expected return equals 0.633%.[3] Using equation (2.10), the estimated theoretical annual expected volatility is 9.511%, and the degree of mean reversion $\beta + \alpha\gamma^2$ is 0.823.[4]

Note that the excepted annual return of 0.633% and expected annual volatility of 9.511% shows that the H-N 2000 parameters lead to rather unusual risk-return levels, compared to more recent data. Although the risk-free rate r is set to 0, because of the log-normal distribution and the existence of the risk premium factor, we should expect higher annual returns. In Figure 2.1 we simulated 4 sample paths in a time horizon of 5000 trading days (or roughly 20 years), and the average stock price is close to the initial price, which implies that the risky asset almost has no excess return. This is quite uncommon for the stock market. Also, the low β and degree of mean reversion means that only a small portion of today's volatility goes into the next day, and this effect vanishes quickly over time. In comparison, the MLE result seen in Christoffersen and Jacobs (2004), Kanniainen, Lin and Yang (2014), and many others who used a longer time series (around 20 years) give a beta of around 0.9 and GARCH persistence close to 1. It should not be surprising that we should aim to fit the model with the relative long return series, because different economical cycles

[3]The theoretical annual expected return is computed using the conditional generating function of S_t. See Heston and Nandi (2000) for the formula and proof.

[4]The approximated annual volatility has a 0.02% absolute difference with the result from Monte Carlo estimation. One error source is that equation (2.10) estimates $\mathbb{E}\left(\sqrt{h(t)}\right)$ by $\sqrt{\mathbb{E}(h(t))}$.

will give different conditional variances that will help identify model parameters, as argued by Christoffersen and Jacobs (2004, online appendix). We now compare these parameter set with C-H-J 2006 parameters.

Christoffersen-Heston-Jacobs 2006 parameters From Christoffersen, Heston and Jacobs (2006), we also obtain the MLE estimates for the parameters using Daily S&P 500 returns from January 3, 1989 to December 30, 2001 for a total of 3324 returns.[5] The estimated parameters are:

$$\begin{cases} \lambda = 2.772 \\ \omega = 3.038 \times 10^{-9} \\ \alpha = 3.660 \times 10^{-6} \\ \beta = 0.9026 \\ \gamma = 128.4 \end{cases} \tag{2.16}$$

We keep the same initial values and zero interest rate

$$r = 0,\ S(0) = 100,\ h(1) = 9.884e\text{-}5, \tag{2.17}$$

and generate 4 different sample trajectories with the parameters in (2.16) and (2.17) and $N = 5000$.[6]

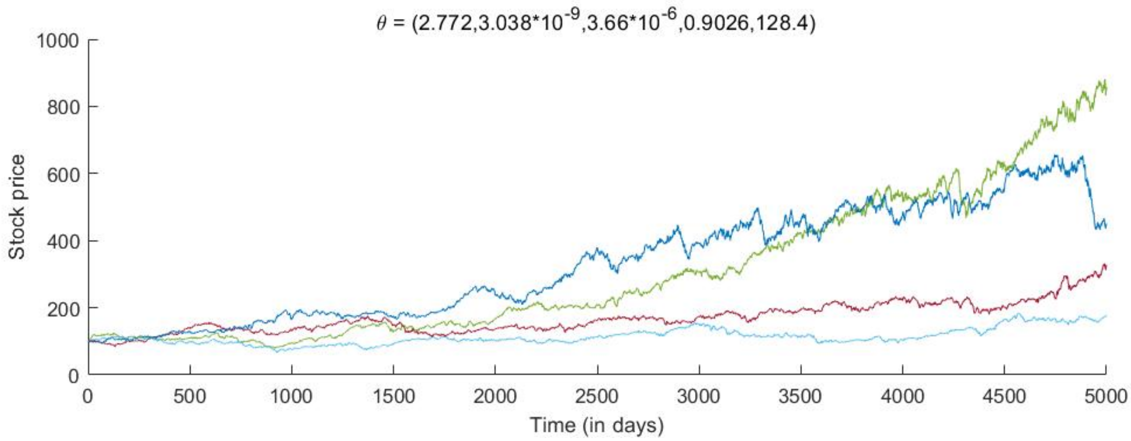


Figure 2.2: Sample trajectories for $N = 5000$, with parameters in (2.16) and (2.17)

[5]The daily total index returns from CRSP are used (see Christoffersen, Heston and Jacobs, 2006). Also, the number 3324 is not from the cited paper, but rather an estimate using the number of trading days during this period.

[6]The value 9.884e-5 is the long-run average conditional variance for parameters in (2.16). The daily disturbances $z(t)$ are exactly the same as the ones in Figure 2.1.

With this set of parameters, the estimated annual expected return equals 8.304%, and the annual expected volatility is 15.626%, both calculated via Monte Carlo simulation with $M = 100,000$ trajectories. The theoretical annual expected return equals 8.271%. Using equation (2.10), the approximated annual expected volatility is 15.782%, and the degree of mean reversion $\beta + \alpha\gamma^2$ is 0.963.[7] From the graph, the 4 simulated paths show a much higher mean excess return. Also, the 16% annual volatility and the persistence of 0.963 are closer to reality, compared to H-N 2000 parameters.

Christoffersen-Heston-Jacobs 2013 parameters From Christoffersen, Heston and Jacobs (2013), we obtain the MLE estimates for the parameters using Daily S&P 500 returns from January 1, 1990 to December 31, 2010 for a total of 5272 returns.[8] The estimated parameters are:

$$\begin{cases} \lambda = 1.094 \\ \omega = 0 \\ \alpha = 3.364 \times 10^{-6} \\ \beta = 0.838 \\ \gamma = 196.82 \end{cases} \tag{2.18}$$

We keep the same initial values and zero interest rate

$$r = 0,\ S(0) = 100,\ h(1) = 1.0617e\text{-}4, \tag{2.19}$$

and generate 4 different sample trajectories with the parameters in (2.18) and (2.19) and $N = 5000$.[9]

[7] The approximated annual volatility has a 0.16% absolute difference with the result from Monte Carlo estimation. One cause is that equation (2.10) estimates $\mathbb{E}\left(\sqrt{h(t)}\right)$ by $\sqrt{\mathbb{E}(h(t))}$.

[8] The MLE is carried such that ω is restricted to 0, so the model only has 4 parameters. The number 5272 is not from the cited paper, but rather an estimate using the number of trading days during this period.

[9] The value $1.0617e$-4 is the long-run average conditional variance for parameters in (2.18). The daily disturbances $z(t)$ are exactly the same as the ones in Figure 2.1 and 2.2.

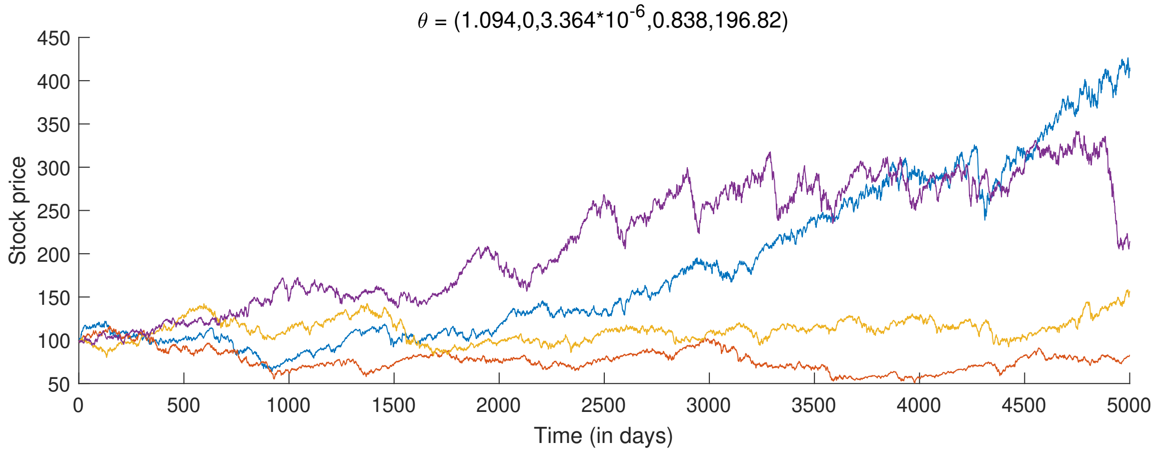


Figure 2.3: Sample trajectories for $N = 5000$, with parameters in (2.18) and (2.19)

With this set of parameters, the estimated annual expected return equals 4.204%, and the annual expected volatility is 15.632%, both calculated via Monte Carlo simulation with $M = 100,000$ trajectories. The theoretical annual expected return equals 4.195%. Using equation (2.10), the approximated annual expected volatility is 16.357%, and the degree of mean reversion $\beta + \alpha\gamma^2$ is 0.968.[10] As a comparison, C-H-J 2013 parameters are more in line with C-H-J 2006 parameters, in terms of values of parameters, implied persistence, and long-run conditional volatility, when compared to H-N 2000 numbers. This is because the return data they use in the parameter estimation has an overlapping period. As a contrast, we see that C-H-J 2013 has a much smaller λ of 1.094, which is slightly more than a third of 2.772, the λ reported in C-H-J 2006. This directly results in a much smaller annual excess return of 4.204%, compared to 8.304% of C-H-J 2006. Moreover, the higher γ implies a higher leverage effect, higher asymmetry, and higher kurtosis compared to C-H-J 2006.

Parameter comparison The following table summarizes the stock price dynamics implied by the choices of parameters

[10]The approximated annual volatility has a 0.725% absolute difference with the result from Monte Carlo estimation. One cause is that equation (2.10) estimates $\mathbb{E}\left(\sqrt{h(t)}\right)$ by $\sqrt{\mathbb{E}(h(t))}$.

Table 2.1: A comparison of different sets of parameters

	λ	ω	α	β	γ	Mean[1]	Ann Vol[2]	Pers	Skew[3]	Kurt[4]
H-N	0.205	5.02e-6	1.32e-6	0.589	421.39	0.63%	9.51%	0.823	0.0048	3.344
C-H-J 06	2.772	3.038e-9	3.66e-6	0.9026	128.4	7.76%	15.78%	0.963	0.0028	3.338
C-H-J 13	1.094	0	3.364e-6	0.838	196.82	4.20%	16.357%	0.968	0.0139	3.889

[1]The (annualized) expected return is estimated via Monte Carlo simulation with $M = 100,000$ paths.

[2]The (annualized) expected volatility is calculated based on (2.10).

[3]This is the sample skewness from 100,000 log returns.

[4]This is the sample kurtosis from 100,000 log returns.

In comparison, the H-N 2000 parameters are estimated with 3 years of data, the C-H-J 2006 parameters are estimated with 13 years of data, while the C-H-J 2013 parameters are estimated with 21 years of data. The respective variance risk premium factor (λ) estimates are: 0.205, 2.772, and 1.094, causing almost no excess return (0.63% for H-N 2000), high degree of excess return (7.76%, for C-H-J 2006), and a moderate amount of excess return (4.20%, for C-H-J 2013). The C-H-J 2006 and C-H-J 2013 have close long-run volatility at around 16%, and persistence. In terms of higher moments, all sets of parameters show positive skews, with C-H-J 2013 having the highest level of skewness, due to large α and γ. Also, the kurtosises of log returns are all close to 3, implying that the conditional distributions of log returns have slightly heavier tails than a normal distribution. We see significant differences when comparing two C-H-J parameters with H-N 2000 ones, especially in annualized expected returns and volatilities, and the set of parameters of C-H-J 2006 and C-H-J 2013 are more consistent with more recent financial data, as seen in the following comparison figure.

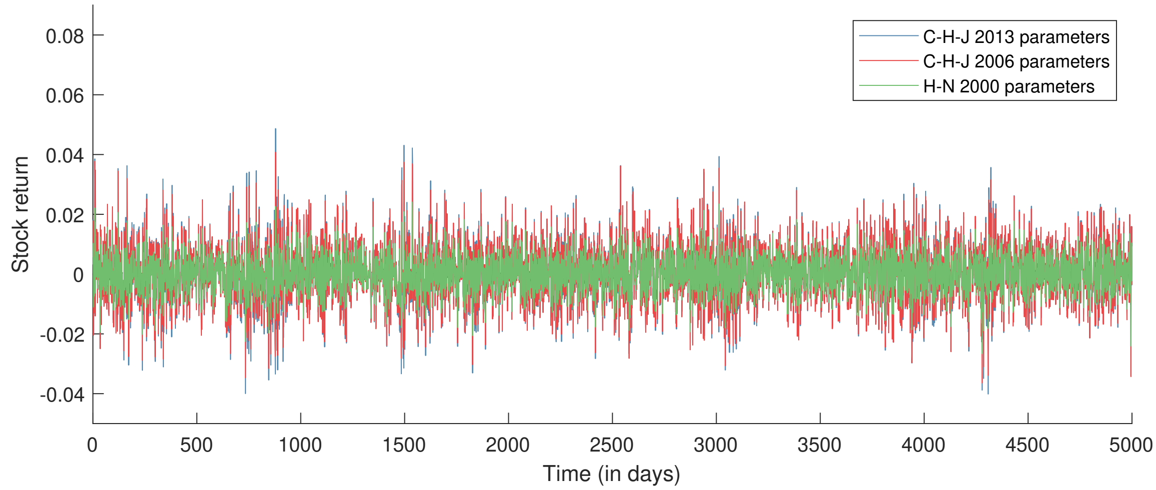


Figure 2.4: A comparison of 5000 log returns (generated using the same disturbances $z(t)$) for 3 sets of parameters

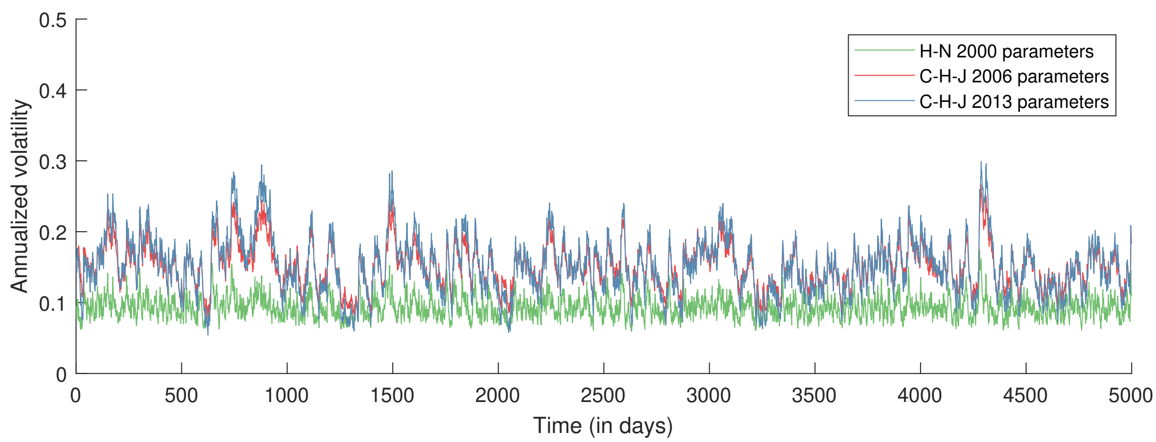


Figure 2.5: A comparison of the daily conditional variance process $h(t)$ (generated using the same disturbances $z(t)$) for 3 sets of parameters[1]

[1]The plot shows the simulated conditional variance process $h(t)$, but with the daily variance transformed into its equivalent annualized volatility via $h(t) \to \sqrt{252h(t)}$.

From both figures, we see that the simulated returns from the two C-H-J parameters are more volatile. In Figure 2.5, we see that the H-N 2000 parameters imply a 10% volatility on average, while the C-H-J parameters both imply an average of volatility at around 16%. Furthermore, the high persistences from both sets of C-H-J parameters are nearly 0.97, which are very close to 1, showing a high correlation of conditional variance between temporally close days, and a relatively long memory of the series $h(t)$. The effect of long memory is clearly shown in Figure 2.4 and Figure 2.5, since at the beginning, the volatility is at a relatively high level, and it takes around 500 trading days for conditional volatility to reach a low level at 10%. Clearly, both C-H-J parameters actively demonstrate a higher mean and variance of the volatility, while the H-N 2000 parameter set is more stable at a lower level.

In conclusion, both the C-H-J 2006 parameters and the C-H-J 2013 parameters are empirically closer to current data and stock behaviours, and are comparable to the estimation results as seen in Kanniainen, Lin, and Yang (2014) and much other recent literature. When comparing C-H-J 2006 parameters with C-H-J 2013, we only see slightly different dynamics as the latter has higher moments for the conditional variances but a substantially lower risk premium. Although the calibration in C-H-J 2013 is done with ω restricted to 0, the difference without this restriction is minimal since the ω estimates are generally very close to 0. Because of numerical optimization challenges, we also use the same restriction of $\omega = 0$ as in Christoffersen, Heston, and Jacobs (2013). As a consequence, we decided to focus mostly on the C-H-J 2013 parameters and regard (2.18) as the true parameters in our simulation study.

2.2.3 Parameter constraints

It is quite important to specify the parametric space, in particular, specifying the range of feasible parameters that we would like to work with. This process is particularly useful in the simulation and calibration process. The parametric space must be specified ahead, especially in the calibration process, as the objective function is highly nonlinear. In this section, we will briefly discuss the issue when the constraints are violated and hence deduce the parameter constraints. In the HN-GARCH model, there are 5 parameters, including λ in the log return process and $\omega, \alpha, \beta, \gamma$ in the volatility process. The ultimate goal here is to find a set of parameters in which the volatility process is strictly positive and stationary with finite variance.

Constrains for λ: Although empirical findings show a positive risk premium, i.e. the level of risk adds value to the asset price, the model does allow λ to be either positive or negative, with an unbounded range. This is because λ only influences log asset return, which could be of an arbitrary amount. Even with a sufficiently negative log return, the asset price still remains positive. Hence, the model does not disapprove a negative λ. As a consequence, there is no constraint for this parameter. We start with the basic assumptions that the stationary GARCH process $h(t)$ is non-negative and has a finite second moment.

Constrains for ω, α and β: In Heston-Nandi (2000)'s original proposal of HN-GARCH, the constraints for the GARCH parameters are not discussed. Hence, we refer back to Bollerslev (1986)'s GARCH setup that all the GARCH parameters, ω, α and β, are non-negative, because we want to avoid the issue that the conditional variance $h(t)$ reaches a non-positive value for any t. Recall equation (2.2)

$$h(t) = \omega + \beta h(t-1) + \alpha \left(z(t-1) - \gamma\sqrt{h(t-1)} \right)^2. \tag{2.20}$$

As we discussed earlier, if $\alpha = 0$, then the process becomes purely deterministic, and if $\alpha = \beta = 0$, then the conditional volatility is constant, determined by ω. Clearly, by assuming that all parameters are non-negative, with a non-negative initial value $h(1)$, and a positive parameter (any of ω, α and β), we can see recursively that the process $h(t)$ will remain positive (if $\omega = \alpha = \beta = 0$, then $h(t) = 0$ for all $t \geq 2$, which is not of interest), however, if any of these restrictions are violated, then the process $h(t)$ might not be positive for some t, for a positive probability, as shown in the following proposition.

Proposition 2.7. *Suppose $h(1) > 0$, and*

$$h(t) = \omega + \beta h(t-1) + \alpha \left(z(t-1) - \gamma\sqrt{h(t-1)} \right)^2$$

for $t \geq 2$. We have the followings:

a) If $\omega, \beta \geq 0$ but $\alpha < 0$, then $\mathbb{P}(h(2) < 0) > 0$.

b) If $\omega, \alpha \geq 0$ but $\beta < 0$, then $\mathbb{P}(h(3) < 0) > 0$.

c) If $\alpha \geq 0$, and $0 < \beta < 1$, but $\omega < 0$, then $\mathbb{P}(h(T) < 0) > 0$ if and only if

$$T > \log_\beta \left(\frac{\omega}{h(1)(\beta - 1) + \omega} \right) + 1. \tag{2.21}$$

Proof. We prove each part individually.

a) The proof is trivial by noting that

$$\mathbb{P}(h(2) < 0) = \mathbb{P}\left(\left(z(1) - \gamma\sqrt{h(1)}\right)^2 > \frac{\omega + \beta h(1)}{-\alpha}\right) > 0.$$

b) We see that if

$$h(2) > \frac{\omega}{-\beta},$$

then

$$h(3) \leq \omega + \beta h(2) < \omega + \beta \frac{\omega}{-\beta} = 0.$$

Hence,

$$\begin{aligned}\mathbb{P}(h(3) < 0) &> \mathbb{P}\left(h(2) > \frac{\omega}{-\beta}\right) \\ &= \mathbb{P}\left(\left(z(1) - \gamma\sqrt{h(1)}\right)^2 > \frac{\omega}{-\beta} - \omega - \beta h(1)\right) > 0.\end{aligned}$$

c) Consider the deterministic recursive sequence $h'(t)$ defined as $h'(1) = h(1)$, and

$$h'(t) = \omega + \beta h'(t-1)$$

for $t \geq 2$. Clearly $h' \leq h$. Now, let T be the smallest integer such that

$$h'(T) < 0.$$

Also, for $\varepsilon > 0$ consider the process $h''(t)$ defined by $h''(1) = h(1)$, and

$$h''(t) = \omega + \varepsilon + \beta h''(t-1).$$

Because as $\varepsilon \to 0$, $h''(T) \to h'(T)$, there must be a sufficiently small $\varepsilon > 0$ such that

$$h''(T) < 0.$$

Now suppose that we have $z(1), ..., z(T-1)$ where

$$\left(z(t-1) - \gamma\sqrt{h(t-1)}\right)^2 < \varepsilon \text{ for } t = 1, 2, ..., T.$$

Clearly, this event has a positive probability to occur. However, with our construction,

$$h(i) < h''(i) \text{ for } i = 1, 2, ..., T,$$

in particular, $h(T) < h''(T) < 0$. Also, since $h''(t) > h(t)$ for all t, if $h'(t) \geq 0$, then $h''(t) \geq 0$. In another word, T is the smallest integer that satisfies

$$\mathbb{P}(h(T) < 0) > 0.$$

Lastly, we provide a formula for finding T. One can easily verify that the process $h'(t)$ defined by the recursive equation has the solution

$$h(t) = \left(h(1) + \frac{\omega}{\beta - 1}\right)\beta^{t-1} - \frac{\omega}{\beta - 1}. \tag{2.22}$$

Hence, by solving for t for the inequality $h(t) < 0$, we get

$$t > \log_\beta\left(\frac{\omega}{h(1)(\beta - 1) + \omega}\right) + 1.$$

$\square$

In conclusion, this proposition gives a very clear message that we cannot have a negative α, β, or ω. Negative parameters could make the conditional variance process $h(t)$ fall below 0, and when that happens, we cannot simulate the process, nor price options. The cases for α and β are very clear, however for ω, we might actually allow a very slight negativity, especially when the time interval T is short, as given by the exact bound (2.21). When we simulate paths with a negative ω, because of the probability for the simulated path to reach a negative $h(t)$ is so small that everything is likely to work just fine, however, due to the fact that option price is merely an expectation under the risk-neutral probabilities, negative ω's will break down the HN-GARCH option pricing formula when T is greater than the bound (2.21), as illustrated below.[11]

Table 2.2: Breakdown T for different parameters

ω	-1e-9	-1e-9	-1e-12	-1e-12
β	0.9	0.95	0.9	0.95
Breakdown T (in days)	110	211	175	347

[11]The initial volatility $h(1) = 1e\text{-}3$, which corresponds to a 50.2% annual volatility. This number is intentionally picked to be larger than usual, however breakdown periods are still short, to emphasize the effect of a negative ω.

It is clear from Table 2.2 that even if ω goes negative by a little amount (in absolute value), the breakdown time T would still be quite short, and Proposition 2.6 would imply that with these chosen parameters, the HN-GARCH analytical formula will fail to perform. As a direct consequence, we cannot have a negative ω. In conclusion, the three GARCH parameters, namely ω, α, and β cannot be negative.

Constraints for γ We see in (2.2) that because the parameter γ is in the squared brackets, whether it is positive or negatives does not influence the negativity of $h(t)$. Therefore, there is no constraints for γ with that regards. However, when γ is too large, the process $h(t)$ may fail to have finite second moment, or fail to stay stationary. Similar to the GARCH model introduced by Bollerslev (1986), we refer to Heston and Nandi (2000) that the persistence of the HN-GARCH is $\beta + \alpha\gamma^2$, as seen if we rewrite the conditional variance into

$$h(t) = (\omega + \alpha) + (\beta + \alpha\gamma^2)h(t-1) + \alpha v(z(t)), \tag{2.23}$$

where $v(z(t+1))$ is a zero mean daily innovation defined by

$$v(z(t)) = \left[\left(z(t) - \gamma\sqrt{h(t)}\right)^2 - (1 + \gamma^2 h(t))\right]. \tag{2.24}$$

Clearly when persistence equals 1, then we have an integrated (non-stationary) variance process since the long-run level of conditional variance is infinity. When persistence is less than 1, then we have a stationary process with a finite second moment (see Heston and Nandi, 2000).

GARCH stationary conditions As a summary, based on above discussions, we restrict our parametric space with the following parameter constraints

$$\text{(C1) - } \begin{cases} \beta + \alpha\gamma^2 < 1 \\ \omega \geq 0 \\ \alpha > 0 \\ \beta > 0 \end{cases} \tag{2.25}$$

Furthermore, with this setting, we have a (covariance) stationary GARCH process, and $h(t) > 0$ for all t.

2.3 Option pricing

With all the propositions with regards to the physical returns of HN-GARCH model, we finally reach the fundamental goal of the model - to efficiently price options. One favorable property about the model is that it provides an almost closed form valuation formula for European options. The standard risk-neutral valuation approach says the price for a European call option with strike price K and time to maturity T is given by

$$c = e^{-rT}\mathbb{E}^*\big((S_T - K)_+\big) \tag{2.26}$$

where $\mathbb{E}^*$ denotes expectations taken with respect to risk-neutral probabilities. The spot asset price process takes the same GARCH(1,1) form in the risk-neutral world, as seen earlier,

$$\log(S(t)) = \log(S(t-1)) + r - \frac{1}{2}h^*(t) + \sqrt{h^*(t)}z^*(t) \tag{2.27}$$

$$h^*(t) = \omega + \beta h^*(t-1) + \alpha\left(z^*(t-1) - \gamma^*\sqrt{h^*(t-1)}\right)^2 \tag{2.28}$$

with λ replaced by $-1/2$ and γ replaced by $\gamma^* = \gamma + \lambda + 1/2$, where $z^*(t)$ has standard normal distribution. With this setup, we present the next proposition regarding European options prices.

Proposition 2.8. *An European call option with strike price K and time-to-maturity T can be computed via*

$$\begin{aligned} C =& \frac{1}{2}x + \frac{e^{-rT}}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi}f^*(i\phi+1)}{i\phi}\right]d\phi \\ &- Ke^{-rT}\left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi}f^*(i\phi)}{i\phi}\right]d\phi\right) \end{aligned} \tag{2.29}$$

where x is the spot price, $\mathrm{Re}[\cdot]$ denote the real component of a complex value, and

$$f^*(\phi) = \mathbb{E}^*\left[S(T)^\phi\right] = S(t)^\phi \exp\left(A(t) + B(t)h(t+1)\right) \tag{2.30}$$

is the generating function of the risk-neutral process (2.27) and (2.28), and the coefficients

$$A(t) = A(t+1) + \phi r + B(t+1)\omega - \frac{1}{2}\log\left(1 - 2\alpha B(t+1)\right) \tag{2.31}$$

$$B(t) = \phi\left(\lambda + \gamma\right) - \frac{1}{2}\gamma^2 + \beta B(t) + \frac{1/2\left(\phi - \gamma\right)^2}{1 - 2\alpha B(t+1)} \tag{2.32}$$

$$A(T) = B(T) = 0 \tag{2.33}$$

can be solved recursively (backwards) with the terminal conditions.

Proof. See Heston and Nandi (2000) Appendix A. Derivation of the Generating Function and Option Formula. □

The reason why the pricing formula (2.29) is called almost closed form is that we must rely on numerical estimation to invert the characteristic function, i.e. to compute the two integrals in equation (2.29). For example, we could use computational software, Matlab, to implement the analytical valuation function that first finds the coefficients $A(t)$ and $B(t)$ recursively, and then numerically estimates the integral (with a default accuracy of $1e$-6) by Matlab's integral function. The integration estimation could get time-consuming, especially leveraged by the number of options we want to calibrate the model with. Thankfully, these integrals converge very fast, both from our experience and as reported by Heston and Nandi (2000).

Besides the uses of an analytical formula, one can alternatively use a Monte Carlo simulation-based approach to numerically generate sample processes from the risk-neutral equations, and estimate the call price C. We now provide a brief comparison to examine the efficiency of our analytical formula. We choose the C-H-J 2013 parameters (see (2.18)) of

$$(\lambda, \omega, \alpha, \beta, \gamma) = (1.094, 0, 3.364e\text{-}6, 0.838, 196.82)$$

and compute both the analytical (true) prices and the MC prices of at-the-money call options with $S_0 = 100$, $r = 0$, and different maturities (1 month, 3 month, 6 month and 1 year). The computational time is the total time for 1000 evaluations. For Monte Carlo price, each price is computed from generating $M = 10,000$ trajectories, and the sample mean and standard deviation are from 1000 prices.

Table 2.3: A comparison of analytical price and MC price of European call options

$S_0 = 100$, $r = 0$, $K = 100$	$T = 23$	$T = 69$	$T = 138$	$T = 276$
Analytical price	1.8999	3.3225	4.7390	6.7531
Computational time (in sec) [1]	2.48	5.65	11.92	33.93
MC price mean	1.8993	3.3220	4.7347	6.7487
MC price standard deviation	0.0254	0.0446	0.0673	0.1033
Computational time (in sec)	92.30	114.66	145.91	241.50

[1]The computational time is the total time for 1000 function calls.

From Table 2.3, we can clearly see that using the analytical formula not only reduces computational time dramatically, but also has an (almost) zero bias. Also, as time-to-maturity T increases, the standard deviation of Monte Carlo prices gets larger, due to the bigger variance of the terminal stock price distribution, making the analytical formula appealing in those situations. Furthermore, since the recursive coefficients $A(t)$ and $B(t)$ in (2.30) is independent of strike prices, spot variance and time-to-maturity, it is even faster to compute a cross-section of options with different strike prices, maturities, and dates under the same GARCH parameters.[12] In conclusion, the benefits of having a closed-form formula over non-affine GARCH option valuation models bring many applications and extensions, including model calibration with option prices, forecasting at a large scale, solving for hedging strategies, and etc.

[12]The recursive sequence of coefficients $A(t), A(t+1), \cdots, A(T)$ and $B(t), B(t+1), \cdots, B(T)$ for short maturities are fully contained in (the pricing formula of) options with longer maturities.

Chapter 3

Maximum Likelihood Estimation of HN-GARCH

The usage of maximum likelihood estimation (MLE) in time series has a rich history since the appearance of the ARCH (see Engle, 1982) and GARCH type models (see Bollerslev, 1986). The well-documented maximum likelihood estimation method, besides its clear intuition, has several attractive properties, including consistency and asymptotic efficiency. Thanks to the development of modern computers and numerical optimization algorithms, once we derive the likelihood function, the process of finding the global maximum parameter merely becomes an optimization problem, that can be solved easily by computational tools, which greatly eases the implementations. In this section, we derive the likelihood function that is used for estimating parameters based on Returns-only MLE, Options-only calibration, as well as Joint estimation-calibration using both returns and options. We discuss the assumptions used, and provide sample-based estimates for the covariance matrix, in order to compute and estimate the standard error of the parameters.

3.1 MLE of return data only

When we only have discrete observations of stock prices, the estimation is very similar to the MLE of the traditional GARCH models (as discussed in Bollerslev, 1986). Suppose we start with some initial time stamp, and based on the initial stock price $S(0)$ we observe the next $N-$day values of stock prices. We wish to establish our maximum likelihood estimates for the parameters

$$\boldsymbol{\theta}' = (\lambda, \omega, \alpha, \beta, \gamma).$$

Let $\boldsymbol{\Theta}$ be a parametric space specified by our parameter constraints (2.25), and assume further that the true parameter is an interior point in $\boldsymbol{\Theta}$ to ensure the termination of the iterative optimization algorithm. With this setup, we can derive the log-likelihood function.

3.1.1 Log likelihood function

Our observation of stock prices $S(0), S(1), ..., S(N)$ are generated from the model, with some true parameter $\boldsymbol{\theta_0}$.[1] In order to find the joint likelihood function, we first transform the stock prices into log prices, namely, $x_0, x_1, ..., x_N$. Because the N-day log stock prices follow some parameter-unknown joint density, denoted by $f_{(X(1),...,X(N))}$, the likelihood function is

$$L(x_1, ..., x_N|\boldsymbol{\theta}) = f_{(X(1),...,X(N))}(x_1, ..., x_N|\boldsymbol{\theta}), \tag{3.1}$$

where $X(i)$ denote the random log stock price on day i. Because of the GARCH specification, we cannot write the joint density explicitly, since the daily log returns are not mutually independent, hence a conditioning trick needs to be applied, so that

$$f_{(X(1),...,X(N))} = f_{X(1)} \times f_{X(2)|X(1)} \times f_{X(3)|X(2),X(1)} \times \cdots \times f_{X(N)|X(N-1),\cdots,X(1)}. \tag{3.2}$$

Since $X(t)$ only depends on $X(t-1)$, and conditioning on $X(t-1)$ and $\boldsymbol{\theta}$, we have (note that $h(t)$ is known given $X(t-1)$ and $\boldsymbol{\theta}$)

$$X(t)|(X(t-1) = x_{t-1}, \boldsymbol{\theta}) \sim N\big(x_{t-1} + r + \lambda h(t), h(t)\big). \tag{3.3}$$

Plugging directly the normal density of (3.3) into (3.2) gives

$$L(x_1, ..., x_N|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi h(i)}} \exp\left(-\frac{(x_i - x_{i-1} - r - \lambda h(i))^2}{2h(i)}\right), \tag{3.4}$$

so the log likelihood function is

$$\begin{aligned}
\ln L^R = l(x_1, ..., x_N|\boldsymbol{\theta}) &= \log L(x_1, ..., x_N|\boldsymbol{\theta}) \\
&= \sum_{i=1}^{N} -\frac{1}{2}\log(2\pi h(i)) - \frac{(x_i - x_{i-1} - r - \lambda h(i))^2}{2h(i)} \\
&= -\frac{1}{2}\sum_{i=1}^{N} \big(\log(2\pi) + \log(h(i)) + z(i)^2\big), \qquad (3.5)
\end{aligned}$$

where the term $z(i)$ is the implied daily noises, given parameters $\boldsymbol{\theta}$ and observations $x_1, ..., x_N$. With the log likelihood derived, the estimated parameters $\hat{\boldsymbol{\theta}}$ is obtained by

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\, l(x_1, ..., x_N|\boldsymbol{\theta}) \tag{3.6}$$

[1] Note that we assume the initial stock price is non-random here. Hence, what we really observe from the stock prices is the daily returns between the $N+1$ days.

within our parametric space. Note that although we need the risk-free interest rate r in computing the log-likelihood function, the parameter r is not estimated directly from stock prices, but rather extracted from other market observables, such as the 3 month Treasury bill rates.

Initial value of $h(t)$ In order to recursively recover the daily noises $z(i)$ and conditional variance $h(i)$ required in the log-likelihood function, one needs the initial values $S(0)$ and $h(1)$. Although stock prices, conditional variance $h(t)$ and daily noises $z(t)$ are all observable given a set of parameters, one needs an initial value for $h(t)$ in order to compute the log-likelihood function. One obvious approach is to add $h(1)$, the initial variance, as a parameter into the optimization process. This is rarely used because of the added complexity in the optimization process. In Heston and Nandi (2000)'s original paper, it is suggested to pick an arbitrary initial value, for example, the long-run average conditional variance implied by the parameters, as the initial variance $h(1)$. They argued that because of the strong mean-reverting structure for the conditional variance process (as discussed in Proposition 2.4), when given a non-accurate initial value $h(1)$, the process will rapidly revert to the "true" process $h(t)$, making the estimation result insensitive to the choice of $h(1)$. We follow Christoffersen et al. (2012) and many others to improve upon this approach by setting up an initial portion of the return series as the "warm-up" period (for example, the first 50 observations), prepared for letting the initial variance to revert to fairly close to the "true" variance. In the warm-up period, the computed variance process $h(t)$ will deviate from the "true" variance, making the calculation of the log-likelihood function inaccurate. As a result, when computing the log-likelihood function, we drop the warm-up period to get a more accurate likelihood value. By doing this, we give up some observations, in exchange for a more accurate initial variance $h(1)$.[2]

3.1.2 Properties of the maximum likelihood estimator

The maximum likelihood estimation of HN-GARCH works similarly as standard ARCH and GARCH time series, and the properties of the estimators are discussed in Bollerslev (1986). Under standard regularity conditions, suppose the observed

[2]There are also other implementations. As an alternative, one can extract information of initial variance from other observable, for example, the VIX series. We refer to Kanniainen, Lin and Yang (2014, section 3.2) for the full discussion of filtering the conditional variances from VIX data.

data is generated from the model with parameter $\boldsymbol{\theta_0}$, we have:[3]

P-1 (Strong consistency): The maximum likelihood estimator converges almost surely to the true underlying parameter

$$\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta_0}. \tag{3.7}$$

P-2 (Functional invariance): If $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator for $\boldsymbol{\theta_0}$, and g is any transformation on $\boldsymbol{\theta}$, then $g(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator for $g(\boldsymbol{\theta_0})$.

P-3 (Asymptotic normality): The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is asymptotically (multivariate) normally distributed with mean $\boldsymbol{\theta_0}$ and covariance matrix $\boldsymbol{I}^{-1}$, that is,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}^{-1}), \tag{3.8}$$

where $\boldsymbol{I}$ is the Fisher information matrix

$$\boldsymbol{I} = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]. \tag{3.9}$$

The implications of these properties are uplifting. First, from strong consistency, we are certain that the bias of all estimated parameters gets small as we increase the sample size. This enhances the confidence in the estimation result, particularly for large samples. The functional invariance property gives information on the maximum likelihood estimates of a function of the HN-GARCH parameters. This is particularly useful, because if we are interested in persistence $(\beta+\alpha\gamma^2)$, long-run variance $\mathbb{E}(h(t))$, or other statistics that depend on the GARCH parameters, we can plug in the MLEs of each parameter, and get the MLEs for these statistics. This property ensures that the MLE of these statistics is transformed value of maximum likelihood parameters. Lastly, the implications of the asymptotic normality are twofold: first, it helps build standard errors and confidence intervals, as we will see shortly; second, it necessarily implies that the maximum likelihood estimator reaches the Cramér–Rao bound, i.e., it is the (asymptotically) unbiased minimum variance estimator (see Cramér, 1999).

[3] The regularity and sufficient conditions are first discussed by Engle (1982) in Section 6 (page 998) where the paper cited the conditions from Crowder (1976). With a similar setup, the properties are further confirmed by Bollerslev (1986) on GARCH models.

3.1.3 Estimate standard error

The asymptotic distribution of our maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, could have been used to construct our standard error estimate, however, the Fisher information matrix is an expectation computed at the true parameter $\boldsymbol{\theta_0}$. Hence, we have two problems: first, we cannot compute the expectation directly, due to the complexity of the log-likelihood function. As a result, we will need to compute the sample estimation (i.e. the averages of the second partial derivatives) as a proxy for the information matrix. Second, we normally do not know the true parameter. Hence, we must replace the Fisher information matrix with its sample analogy, namely, the observed information

$$\boldsymbol{J}\left(\boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} l\left(x_i|\boldsymbol{\theta}\right), \tag{3.10}$$

where the matrix is often evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. Many have stated the preference of using observed information matrix instead of the Fisher information for finite samples (see Efron and Hinkley (1978) for a theoretical justification and Maldonado and Greenland (1994) provides a simulation-based study that also provides evidence). We have seen that $\hat{\boldsymbol{\theta}}$ is a strongly consistent estimator of $\boldsymbol{\theta_0}$, and the fact that $\boldsymbol{J}\left(\boldsymbol{\theta}\right)$ is a strongly consistent estimator of $\boldsymbol{I}\left(\boldsymbol{\theta}\right)$ for any parameter $\boldsymbol{\theta}$ just follows from the strong law of large numbers (SLLN). However, it is also true that $J(\hat{\boldsymbol{\theta}})$ (the observed information evaluated at finite sample MLE) is a consistent estimator for Fisher information $I(\boldsymbol{\theta})$. Note that because $J(\boldsymbol{\theta})$ is also a function of N, so we cannot apply continuous mapping theorem, but need to rely on some equicontinuity conditions for the class of functions J_N. The proof can be found online (see Dandar, 2013), and it is made possible thanks to the strong consistency of our maximum likelihood estimator. Now, based on this fact, the asymptotic properties of the maximum likelihood estimator and the Slutsky's Theorem, we can write

$$\hat{\boldsymbol{\theta}} \xrightarrow{d} N(\boldsymbol{\theta_0}, \frac{1}{N}\boldsymbol{J}^{-1}) = N(\boldsymbol{\theta_0}, (N\boldsymbol{J})^{-1}). \tag{3.11}$$

However, by Information Matrix Equality, the Hessian matrix equals the outer product of the gradient (see Bollerslev (1986) for GARCH models, and Papadopoulos (2004) for regularity conditions). As a direct consequence, we can compute the outer product of the gradient for observations $x_1, ..., x_N$, via

$$\hat{\boldsymbol{V}}^{-1} = \sum_{i=1}^{N} \left(\frac{\partial}{\partial \boldsymbol{\theta}} l(x_i|\hat{\boldsymbol{\theta}})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}'} l(x_i|\hat{\boldsymbol{\theta}})\right), \tag{3.12}$$

where $\hat{\boldsymbol{V}}$ is our sample based consistent estimator for the variance covariance matrix. The estimated standard error can be obtained by taking the square root of the diagonal elements of $\hat{\boldsymbol{V}}$.

Simulation based standard error Alternatively, if we have the luxury of simulating samples based on a known true parameter, we can also estimate the MLE standard errors by simulation M trajectories, each having N observations (N days), and hence we can obtain M different MLE estimates. We calculate the sample standard deviation of the M estimates as simulation-based standard error. By this simulation-based standard error, we have an idea of the variability and efficiency of our maximum likelihood estimators, when the model is correctly specified. Of course, when M and N are sufficiently large, the simulation-based standard errors should be close to Fisher-based standard errors (computed from any sample).

Confidence interval The estimated standard error from either approach can help construct confidence intervals. Utilizing the asymptotic properties of our maximum likelihood estimator, an (asymptotic) $100(1-\alpha)\%$ confidence interval for parameter θ_i can be constructed by

$$\hat{\theta}_i \pm z_{1-\alpha/2} \times \hat{\sigma}_i(\hat{\theta}_i), \tag{3.13}$$

where $\hat{\sigma_i}$ is the estimated standard error for $\hat{\theta}_i$.

3.2 MLE of option data only

It is certainly believed that in addition to return data, observed option prices often contain valuable information that can be used to calibrate our option pricing model. This is particularly helpful when our HN-GARCH model attains a closed-form option valuation formula that allows efficient computation of option prices at a large scale. Assuming that the model is correctly specified, a natural approach in fitting the model is to find the parameters that match the observed/market option prices with the model theoretical prices through the minimization of some loss function. The uses of loss functions in calibration with option prices are used in much of the existing literature, including Bakshi, Cao, and Chen (1997), Heston and Nandi (2000), Christoffersen and Jacobs (2004) and many others. We now outline some of the popular loss functions in actual usage.

3.2.1 Loss functions

Heston and Nandi (2000) originally considered a nonlinear least square approach, that is essentially minimizing the sum of squared dollar errors, or the absolute MSE

$$\$ \text{ MSE } = \frac{1}{N}\sum_{i=1}^{N}(C_i - C_i(\boldsymbol{\theta}))^2, \tag{3.14}$$

where C_i and $C_i(\boldsymbol{\theta})$ are respectively the observed (or market) option price and the theoretical model price of the i^{th} option. We find model parameters $\boldsymbol{\theta}$ by minimizing this loss function. From this approach, an apparent disadvantage is that the options with higher price weight more than options with lower price, which makes deep out-of-the-money options, that could contain valuable information with regards to specific parameters, less significant during the calibration process.

To incorporate options with low prices into the fitting process, we could also consider minimizing the relative MSE

$$\% \text{ MSE } = \frac{1}{N}\sum_{i=1}^{N}((C_i - C_i(\boldsymbol{\theta}))/C_i)^2. \tag{3.15}$$

Note that this loss function can be regarded as a weighted MSE where the weight is the reciprocal of the observed option price. However, by doing this, the extremely cheap options are given abnormally large weights in estimation.

As Christoffersen (2004) pointed out, because of the above scale considerations, and also because of the historical convention of quoting option prices by its implied volatility, we can use a implied volatility MSE

$$\text{IVMSE } = \frac{1}{N}\sum_{i=1}^{N}(\sigma_i - \sigma_i(\boldsymbol{\theta}))^2 \tag{3.16}$$

as objective function, where the implied volatilities σ_i and $\sigma_i(\boldsymbol{\theta})$ are computed via inverting the Black-Scholes formula

$$\sigma_i = BS^{-1}\left(C_i, T_i, K_i, S_t, r\right), \text{ and } \sigma_i(\boldsymbol{\theta}) = BS^{-1}\left(C_i(\boldsymbol{\theta}), T_i, K_i, S_t, r\right). \tag{3.17}$$

Despite that the IVMSE metric provides intuitive weightings across all strike prices and maturities, and some other advantages as argued by Broadie et al. (2007), it is quite computationally intensive, as the evaluation of our objective function involves

$2N$ numerical inversions of the Black-Scholes formula, which is computed from pure numerical algorithms. As an alternative, we can approximate the option prices as a linear function of implied volatility times vega

$$\Delta C = \Delta\sigma \frac{\partial C}{\partial \sigma} \tag{3.18}$$

and write out the vega weighted loss function, defined as

$$\text{(Vega weighted) MSE} = \frac{1}{N}\sum_{i=1}^{N}((C_i - C_i(\boldsymbol{\theta}))/\nu_i)^2, \tag{3.19}$$

where ν_i is the Black-Scholes vega calculated at implied volatility by market option price C_i. Note that the proposal of using vega weighted (VWMSE) is merely a numerical trick to overcome the computational burden of IVMSE. In this thesis, we follow Christoffersen et al. (2012) and others to focus on using vega weighted MSE to calibrate our model.

3.2.2 Maximum likelihood-based inference

With Assumption 2.3, given a set of parameters, we can write out the likelihood function of observing $e_1, ..., e_N$

$$L^O = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{e_i^2}{2\sigma^2}\}, \tag{3.20}$$

so the log likelihood equals

$$\ln L^O = -\frac{1}{2}\sum_{i=1}^{N} \ln(2\pi) + \ln(\sigma^2) + \frac{e_i^2}{\sigma^2}, \tag{3.21}$$

where the true standard deviation σ is estimated by the vega weighted root mean squared error (VWRMSE)

$$VWRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((C_i - C_i(\boldsymbol{\theta})/\nu_i)^2}, \tag{3.22}$$

i.e., the square root of sample second moment, since vega error is assumed to have a mean of 0. As a result, we can find the parameters $\boldsymbol{\theta}$ that maximizes the option log likelihood function

$$\ln L^O \propto -\frac{1}{2}\sum_{i=1}^{N}\left\{\ln\left(VWRMSE^2\right) + e_i^2/VWRMSE^2\right\}. \tag{3.23}$$

3.2.3 Conditional variances

The above estimation procedures seem complete, however, we need to note that, as seen in (2.29), the option price on day t is not only a function of option specifications and HN-GARCH parameters, but also the conditional variance $h(t)$. As a consequence, if we have observed N option prices that come from k different days, then we also need to add the conditional variances $h(t_1), ..., h(t_k)$ into the optimization arguments. As a consequence, the total number of parameters to be estimate is $k + 5$. In reality, this approach is very rarely used, because of the added optimization complexity from the additional arguments. Instead, the popular approach relies on filtering the volatility series from existing market information. We introduced earlier two methods of parameter calibration with option prices only, and the only difference is that method 1 extracts the daily variance information from other market data, and is kept fixed throughout the optimization process. As a comparison, method 2 filters the daily variances from the returns, therefore the variances are updated at each iteration of the optimization. As noted by Christoffersen and Jacobs (2004), although method 2 does not compute the return likelihood, it still requires daily return observations in order to filter out the variances. Based on the online appendix of Christoffersen and Jacobs (2004), we can write out the volatility updating equation

$$h(t+1) = \omega + \beta h(t) + \alpha \frac{\left(\log(S(t)) - \log(S(t-1)) - r - \lambda h(t) - \gamma h(t)\right)^2}{h(t)}, \quad (3.24)$$

and so given the asset returns, the initial variance and GARCH parameters, we obtain the filtered variance series.

To our knowledge, the Options-only calibration method 1 is very rarely used. However, we can take advantage of our simulation studies, and pretend we know the true variances, to compare the two calibration methods with options only. In the comparison, we directly give method 1 the true variances, which are then kept fixed throughout the numerical optimization process. Since method 2 needs to filter $h(t)$ from returns, we expect that it is superior to method 1, since filtered variances bring more information. On the other hand, the filtered variances will never be exactly the true ones, hence it is not clear beforehand which method gives a better calibration result. Thus, the comparison between the two methods is motivated.

3.2.4 Properties of MLE estimator

Unlike return MLE, since option prices are only a function of $\gamma^* = \gamma + \lambda + 1/2$, we could not distinguish between γ and λ in option only calibration.[4] This leaves us with only 4 HN-GARCH parameters, compared to 5 with Returns-only estimation. Also, since the observations of different type of options are not identically distributed, we can only refer to Hoadley (1971) for additional conditions required to obtain the asymptotic properties for maximum likelihood estimation for the independent but non-identical case.[5] It is theoretically hard to check with the conditions proposed by Hoadley (1971). However, from a standard non-linear regression point of view, standard properties of MLE should still hold. Thus, we make the assumptions that all the regularity assumptions hold, to finally get the standard error of estimators and confidence intervals from the same formulas as in (3.12) and (3.13). Later on, we will check the consistency of MLE estimators and their standard error estimates by simulated samples.

3.3 MLE of return and option data jointly

In the previous sections, we have seen approaches that estimate model parameters from returns or options alone. However, historical returns only contain past information, while option prices are forward-looking. In addition, more observations will indeed improve the accuracy of the estimation. Because of this, we follow Christoffersen et al. (2012) to provide a framework for the joint estimation using return and option data. Note that the likelihood function L^R for return data alone is derived in (3.5), and the likelihood function L^O for option data when assuming a normal distribution for the vega weighted error is in (3.23). Since the daily disturbances of stock returns and the option vega errors are assumed to be independent, we can add the two-part log-likelihood functions to form a joint likelihood $\ln L^J$

$$\ln L^J = \ln L^R + \ln L^O. \tag{3.25}$$

In equation (3.25), the idea of adding the single part likelihood functions is merely from the hypothetical expectation that the noises in daily stock prices are independent of the option errors. Nonetheless, this does not restrict the option error

[4]In (3.24), the terms $\lambda h(t) - \gamma h(t)$ imply that λ is not identified from $\lambda + \gamma$ from filtered conditional variances.

[5]Note that option prices depend on strike prices K, time to maturity T and conditional variance $h(t)$. With different values of K, T and $h(t)$, the distributions of observed option prices (with vega errors) C_i are different, making the observations non identical.

structure or the distributional assumption at all. If we wish, we could use other structural and distributional specifications of the option errors and replace the vega weighted Gaussian likelihood $\ln L^O$ with other likelihood functions to suit different hypotheses.

In writing such log likelihood function, although the calculation of option likelihood requires conditional variances, such conditional variance process is already filtered out from the return likelihood computations. As a result, the joint estimation contains the conditional variances, and so the estimated parameters are $\boldsymbol{\theta} = (\lambda, \omega, \alpha, \beta, \gamma)'$, which can be obtained from solving the optimization problem

$$\hat{\boldsymbol{\theta}} = \mathrm{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ln L^J(\boldsymbol{\theta}). \tag{3.26}$$

Likewise, because the observations in this joint maximum likelihood estimation are independent but not identically distributed, in particular, the Returns-only MLE has 5 parameters to be estimated, while the Options-only calibration only has 4, we refer to Hoadley (1971), page 1983, for more conditions that are required to obtain asymptotic properties.[6] Under the non-linear regression context, we can positively anticipate consistency to hold in an unusual way. When the number of return observations in the joint sample tends to infinity, the MLE estimates will converge to the true parameters. However, if the number of option observations tends to infinity in a joint sample, while the number of returns is bounded, then only the risk-neutral GARCH parameters α, β, ω, and $\gamma^* = \gamma + \lambda + 1/2$ will converge to the true value. In terms of the Fisher information-based standard errors for our estimated parameters, the formula estimates the covariance matrix by inverting the observed information matrix. Since the return and option observations have different log-likelihoods, the original equation (3.12) should be changed to

$$\hat{\boldsymbol{V}}^{-1} = \sum_{i=1}^{N} \left(\frac{\partial}{\partial \boldsymbol{\theta}} l_i(x_i | \hat{\boldsymbol{\theta}}) \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}'} l_i(x_i | \hat{\boldsymbol{\theta}}) \right), \tag{3.27}$$

where l_i is the log-likelihood observing the ith data. Since for an independent sample, the Fisher information for the whole sample is the sum of single Fisher information, regardless of the underlying distribution, we anticipate the estimating formula (3.27), to be consistent as well.

[6] Here we only list the sufficient conditions if one wants to derive theoretically the asymptotic properties of this joint estimation. However, considering the difficulties to directly check these conditions, we show the results from simulation-based studies that provide a practical, empirical confirmation of some of the properties.

Chapter 4

Empirical Results

In chapter 3, we saw that once we derived the log-likelihood function for fitting return data, and the loss function for option data, the maximum likelihood estimation is merely an optimization problem. In this chapter, we discuss and analyze the model estimation and calibration results when fitting our simulated data. In this chapter, all the optimization is completed using Matlab's solver `fminsearch` that utilizes the derivative-free Simplex search algorithm for doing all the optimization. We will focus on discussing the main results of optimization and their implications in this chapter, and leave the details, including some other optimization results, and numerical settings in the appendix.

4.1 Data simulation

While return data can be easily simulated through the recursive GARCH equations, we still need to simulate option data with the aim to investigate the empirical properties of the joint maximum likelihood estimation. The option data simulation consists of two steps. First, because HN-GARCH attains a closed-form option valuation formula, we focus on determining the combination of strike prices K, time-to-maturities T and conditional variances $h(t)$, that creates a representative sample of option observations. Second, we need to impose noises on true option prices. If we only use true price to calibrate the model, then 4 option prices are enough to pin down all parameters, with any objective function (because the model has 4 identifiable parameters). Hence, the structure of noise is also important. Because of put-call parity, we can easily transform out-of-money call options into in-the-money put options, and vice versa. Therefore, in our studies, we only focus on simulating and using call options.

Simulation of true prices In determining the combination of K, T and h, a natural approach is to use a grid of evenly spread combinations, so that the option sample does not lose characteristics from certain types of options. As an example, we could use a combination of

$$\begin{cases} K = 95, 100, 105, 110, 115, \\ T = 23, 46, 69, ..., 276, \\ h(t) = 0.8e\text{-}5, 0.82e\text{-}5, 0.84e\text{-}5..., 1.18e\text{-}5, \end{cases} \tag{4.1}$$

and calculate the true price for every K, T and $h(t)$. This gives a sample of $5 \times 12 \times 20 = 1200$ options. Doing this grid approach has several benefits. First, our sample contains options with any strike price, and maturity and any volatility. Second, we can always change how fine the grid should be for certain K, T or $h(t)$, to meet different needs. Also, we can adjust the sample size by generating each option twice, or 3 times, resulting in a larger sample. Third, the distribution of different types of options is relatively even. In contrast, a real option sample of call options will contain mostly short-term (less than 60 days) options, and fewer long-term (more than 180 days) options. Moreover, we indeed expect that for deep out-of-money call options, one with long maturity will be more liquid than one with a short maturity. Also, for in-the-money call options, there will be more with short maturity than long maturity, hence creating an imbalance, compared to our simulated sample.

In our simulation study, we will obey this grid approach in simulating option prices. However, when we calibrate on option prices using Option-Returns-NLS, we would use the filtered conditional variance from the return series into computing model option price. Because of this, it also seems plausible to use the conditional variances from the return series, instead of the grid. If we were not using simulated option prices, but rather the market observed prices, a well-documented approach is to use only the Wednesday option prices, because they are less likely to be affected by day-of-the-week effects (see Christoffersen, Jacobs and Mimouni (2010)). As a consequence, once we have the return series generated, we also have the true conditional variance series $h(t)$. Hence, we can use the days with a 5-day difference, say $h(1), h(6), h(11), ...$ to generate true option prices (the strike price K and time-to-maturity T also follow the grid in (4.1)) which are understood as the true option prices on Wednesdays. Because the stationary distribution of variances $h(t)$ is close to a shifted Chi-squared distribution, we expect the option sample will mimic the reality in terms of the economical cycles, where the grid approach uses uniformly distributed variances.

Noise structure Recall that we have assumed that the observed option price equals true price plus a vega weighted noise

$$e_i = (C_i - C_i(\boldsymbol{\theta}))/\nu_i, \tag{4.2}$$

which is assumed to follow a $N(0, \sigma^2)$ distribution. Hence, we can generate random errors $e_1, \ldots,$ to generate true prices. Note that ν_i is the vega calculated by the market price implied volatility, because the true price is unknown. Here we do (and only) know the true price. Hence, in generating such sample, we need to compute the implied volatility and option vega at true prices, and so the sample option price (with noise) is

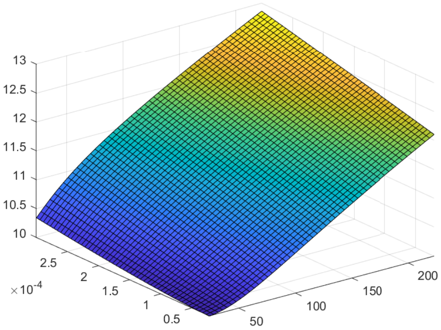
$$C_i = C_i(\boldsymbol{\theta}) + \nu_i e_i, \tag{4.3}$$

For the choice of σ (the standard deviation of option vega noises), we refer to Kanniainen, Lin and Yang (2014) that reported the estimation result (see table 1 and table 4) on the "Options-Returns-NLS" volatility extraction approach using Wednesdays option prices from January 1996 through December 2009, with a 4.9591% vega weighted RMSE.[1] Therefore, we will use this number as a benchmark.

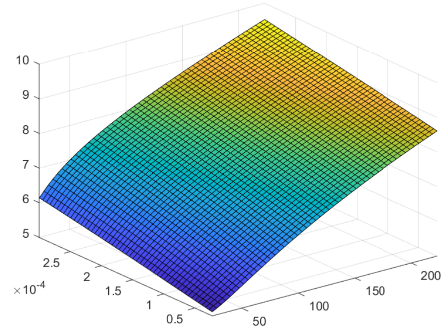
A particular fact in generating option samples based on the vega errors is that when the true option price is close to 0, equation (4.3) might give a negative sample price. The probability of such scenarios is very low, and in such cases, we forget the financial meaning behind option prices, and think of the fitting process as a regression. Hence, a possible drawback is that because the calculation of the vega error (as in (4.2)) requires the calculation of the implied volatility and option vega at market price, whereas the inverse of the Black-Scholes formula cannot take a negative price as an argument. Luckily, we know the true vega (i.e. the vega computed at the true option price) of the option, hence we could use the true vega instead of the market vega in the numerical optimization process, which again alleviates the problem. When the market price is close to the model price, by continuity the implied volatility and vega should also be close. Hence, replacing the vega computed at market prices by the true vega should not harm our results, considering this is a weighted least squares approach.

Statistics of option sample First, we refer to Figure 4.1 for a surface of option prices for each strike price K, where the x- and y-axis are the time to maturity T and conditional variance $h(t)$, and the parameters are kept fixed using (2.16). It is

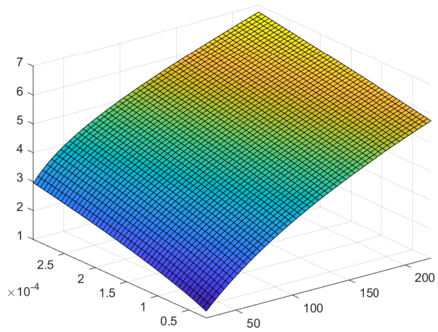
[1]There are 28,096 call options and 40,432 put options in the sample. Also, the number 4.9591% is the sample estimate of σ.



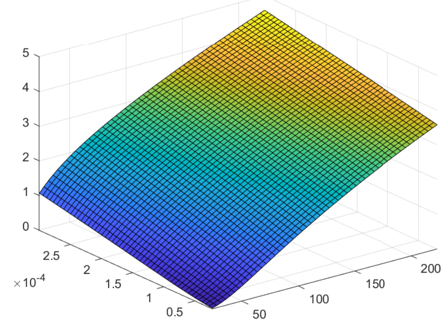
(a) $K = 90$



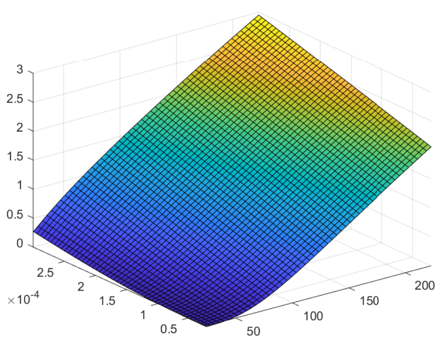
(b) $K = 95$



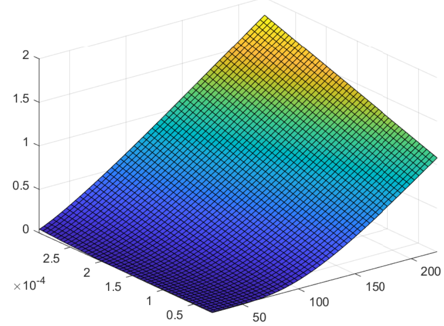
(c) $K = 100$



(d) $K = 105$



(e) $K = 110$



(f) $K = 115$

Figure 4.1: Option prices for different K, T and $h(t)$[1]

[1]The x-axis represents the time-to-maturity T (in trading days), the y-axis is the daily conditional variance $h(t)$, and the z-axis is the option price.

clear that option prices are monotonic in all of K, T and $h(t)$. This is consistent with the fundamental Black-Scholes option pricing framework.

Next, we generated our option sample set by first computing true prices with

$$\begin{cases} K = 95, 100, 105, 110, 115, \\ T = 23, 46 \\ h(5), h(10), ..., h(250), \end{cases} \tag{4.4}$$

where the $h(t)$ are 50 conditional variance from the return series, for a total of 500 options. We present a histogram to show the empirical distribution of the conditional variances $h(t)$.

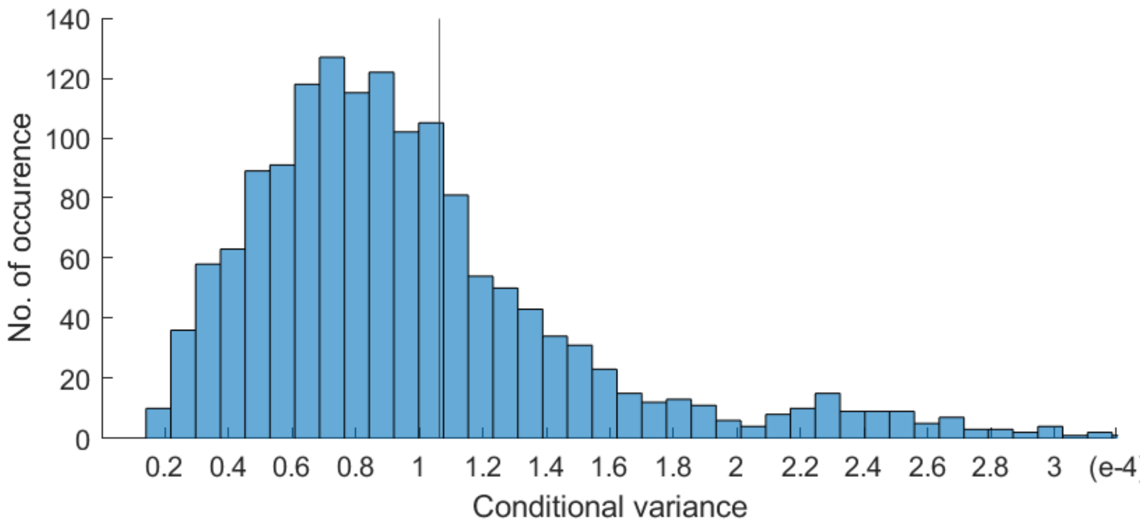


Figure 4.2: Empirical distribution (histogram) of weekly conditional variances $h(t)$

From this figure, we can see that the empirical daily conditional variances show the shape of a non-central Chi-square distribution. Also, although the average level is 1.062e-4, as shown in the graph from the vertical line, the daily variances have a large variation, from 0.2e-4 to 3.2e-4. More importantly, if the variances with 5-day lag are drawn from the sample, we would expect most of the daily variances to be below the long-run level, while a small number of daily variances could have up to triple the size of long-run variance.

Next, we generate random errors e_i from a zero-mean i.i.d. normal random variables with $\sigma = 4.9591\%$, and use (4.3) to compute the option prices in our sample. Our options have 5 categories of strike prices and 2 categories of time to maturities, and the average option price (with noise) and the number of options (in parenthesis) are reported in the table.

Table 4.1: Basic statistics of simulated option sample

Strike price	Time-to-maturity		Total
	1M	2M	
$K = 95$	5.340 (50)	6.378 (50)	5.859 (100)
$K = 100$	1.796 (50)	3.210 (50)	2.503 (100)
$K = 105$	0.190 (50)	1.158 (50)	0.674 (100)
$K = 110$	0.005 (50)	0.206 (50)	0.105 (100)
$K = 115$	0.0003 (50)	0.028 (50)	0.0142 (100)
Total	1.466 (250)	2.196 (250)	1.831 (500)

We note that there are more out-of-money call options than in-the-money ones, because in general investors favour out-of-money call options due to their high leverage and cheap price. We see from Kanniainen, Lin and Yang (2014, table 1) that in-the-money call options, with moneyness (defined as S/K) greater than 1.03 only account for 8.5% of the total option sample. In our sample this number is 20%, and the rest are at-the-money and out-of-money call options, therefore our option sample reflects this structure. Also, the deep out-of-money options may help to estimate certain parameters, because of their high relevance to extreme events. As to our simulation study, we can take advantage and increase the proportion of deep out-of-money call options, as seen in our sample. In terms of time-to-maturity, we only have 1-month and 2-month options to reduce the computational burden. We suspect the main conclusion does not change whether we add options with longer maturity to the sample.

4.2 Returns-only estimation

We first showcase the results under Returns-only estimation, where we apply maximum likelihood estimation on simulated return data, and we analyze the MLE results. Using the C-H-J 2013 parameters, we can generate our return series. We use Matlab's numerical optimizer `fminsearch` to solve for the MLE parameters that maximize the return likelihood function (3.6). In fact, from our experience, the numerical optimization under Returns-only estimation is stable, since `fminsearch` can always converge to the same value with different initial values. This actively confirms our estimation results. The complete result is reported in Table 4.2 where the return series for various N is generated, and the estimates are included. Note that the maximum likelihood estimates reported are for one sample, where the sample standard error is the sample standard deviation for 100 simulated samples. The standard errors estimated via the outer product of gradients are shown in parenthesis. In addition, we also report the persistence, defined as $\beta + \alpha\gamma^2$, and annualized volatility (estimated via (2.10)). As we discussed in section 3.1.1, the estimation result is insensitive to the choice of initial variance. Therefore, in the numerical estimation, we will assume the true initial variance $h(1)$ is known and focus more on the empirical demonstrations on the MLE properties.

The maximum likelihood estimates with $N = 4500$ are close to true sizes with available data, since the size of the sample (around 17.9 years of data) is in line with typical empirical estimations. We see first that parameter λ has an estimated standard error of 1.473, and a sample standard error of 1.335. With the (maximum likelihood) estimated value of 0.902, the size of estimated and true λ is smaller than its standard error, implying that we cannot reject the hypothesis of $\lambda = 0$, and the estimate of λ with $N = 4500$ return data is statistically insignificant. As a comparison, for the GARCH parameters α, γ, both the estimated and sample standard errors are around a tenth of the true parameter value, and for β, the standard error is two digits smaller than the true value. As a consequence, with a great probability, the maximum likelihood estimate lies within $\pm 15\%$ of the true values for α and γ, and $\pm 3\%$ for β. From the table, we see that the accuracy of the estimates is decent, in terms of absolute error, except for λ. However, we have to note that the point estimation result is highly dependent on the particular sample (random seed), hence the standard errors should contain more information in evaluating the efficiency of the estimator.

If we compare horizontally the estimated parameters with different N, we can indeed see a decrease in the difference between estimates and true parameter values.

Table 4.2: Returns-only MLE result with various N

Parameter	True value	$N = 4500$		$N = 45,000$		$N = 450,000$	
		MLE (SE)	Sample SE	MLE (SE)	Sample SE	MLE (SE)	Sample SE
λ	1.094	0.902 (1.473)	1.335	1.579 (0.463)	0.460	1.154 (0.145)	0.138
ω	0	0 (0)	0	0 (0)	0	0 (0)	0
α	3.364e-6	3.756e-6 (4.135e-7)	3.297e-7	3.254e-6 (1.067e-7)	1.002e-7	3.325e-6 (3.434e-8)	3.086e-8
β	0.838	0.845 (1.296e-2)	1.217e-2	0.838 (4.015e-3)	3.447e-3	0.837 (1.270e-3)	1.120e-3
γ	196.82	177.93 (15.792)	16.537	200.51 (5.148)	4.951	199.24 (1.612)	1.216
$\beta + \alpha\gamma^2$	0.968	0.973	3.578e-3	0.969	1.044e-3	0.969	0.294e-3
Annualized Vol.	16.36%	17.64%	0.614%	16.16%	0.195%	16.34%	0.057%

The above table shows the maximum likelihood estimation result with returns only. The data is simulated based on the true parameters (C-H-J 2013 parameters, see (2.18) and (2.19)). The estimated parameters for one sample with various N are shown, together with the standard error computed via the outer product of the gradient shown in parenthesis. The sample standard error is computed by sample standard deviation of the MLEs for 100 simulated samples for each N. The persistence ($\beta + \alpha\gamma^2$) and annualized volatility (computed via (2.10)) for true parameters and MLEs are also included, together with their sample standard deviation.

The comparison is more obvious if we compare the $N = 4500$ estimates with those from $N = 450000$. Notice that $N = 45000$ will correspond to around 179 years of data, which is impractical. Therefore, this exercise is to only empirically verify the consistency of our MLE parameters. From the table, even with a very long series, we can pin down parameters α, β and γ, but the accuracy for λ is still not favourable, proving that the risk premium factor is certainly hard to estimate. Nevertheless, the $N = 4500$ MLE parameters imply the persistence and annualized volatility, and they are both close to the true values, because the persistence does not involve λ, and the annualized volatility will highly match the volatility from the return series. In terms of the estimated standard errors, by comparing horizontally we can see a clear convergence as well. Note that the standard error estimates have errors from 2 sources: the Fisher information estimated by observed information (an average of our N observations), and the replacement of true parameters by the MLEs. Hence, for finite samples, because the MLEs may have a huge deviation from the true parameters, we might see standard error estimates that are not so accurate, and the sample standard errors are more reliable. When fitting real data, although we have to stick with the estimated standard errors, we do not require a high level of accuracy because it only provides a broad indication for our maximum likelihood estimates.

As a summary, when the HN-GARCH model is the true data generation process, we have the consistency of our maximum likelihood estimates, as well as our standard errors estimated by the outer product of the gradient of the log-likelihood function. When comparing the size of standard errors with true parameters, at the practical level of $N = 4500$, the maximum likelihood estimates for α and γ will lie within $\pm 15\%$ region of the true parameters, and $\pm 3\%$ for β, implying a 15% (3% for β) the relative estimation error. In addition, the estimated annualized volatility and persistence are very accurate. From these, we conclude that the Returns-only estimation can capture GARCH parameters, as well as persistence and long-run variance with good accuracy. However, the estimation for λ is not significant, as the size of standard error exceeds the size of the true λ value.

Moreover, the estimated standard errors do not show a large deviation (the deviation varies from 0% to 20%), when compared with sample standard errors. When we compare the standard error estimates with sample standard errors at $N = 4500$ level, we see the numbers are very close for λ, β and γ. Also, if we compare horizontally, we see that as N increases, the difference between standard error estimates and sample standard errors tends to decrease. This confirms the fact that the standard error estimates are asymptotically unbiased, and so we conclude that the gradient-based estimation of standard errors provides a reliable reference to the unknown true standard errors.

4.2.1 Estimation of λ

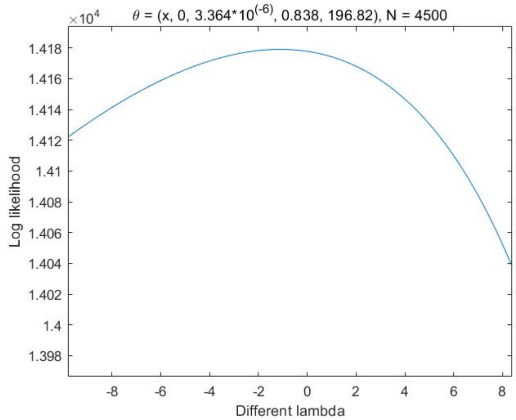
Because of the financial intuitions behind the price of risk parameter λ, we explore the factors behind the difficulties of estimating λ in the Returns-only MLE. The first evidence appears when we observe that the log-likelihood is very insensitive to change in λ. We plot the partial log-likelihood functions with respect to each parameter, in Figure 4.3, where all other parameters are kept fixed at the true parameter level. From the figure, we see in (a) that a change from $\lambda = 0$ to $\lambda = 5$ only decreases log-likelihood by 40. As a comparison, changing β from 0.84 (close to the global maximum) to 0.82 decreases the log-likelihood function by 100. In fact, all other parameters (besides λ) have much bigger impacts on the log-likelihood function compared to λ, making λ very hard to estimate. This also explains the very large standard errors and confidence intervals for λ estimates, as they are computed by computing the first-order partial derivatives with respect to the log-likelihood function.

Another observation that explains why λ estimates are very inaccurate for small samples is that the particular random seed has a large impact on the λ estimate. Although this is true for all parameters, we are particularly talking about the very strong correlation between λ estimates and the mean of daily noises for the simulated sample. Figure 4.4 is a scatter plot of the mean of noise versus λ estimate, from where the correlation is very clear.

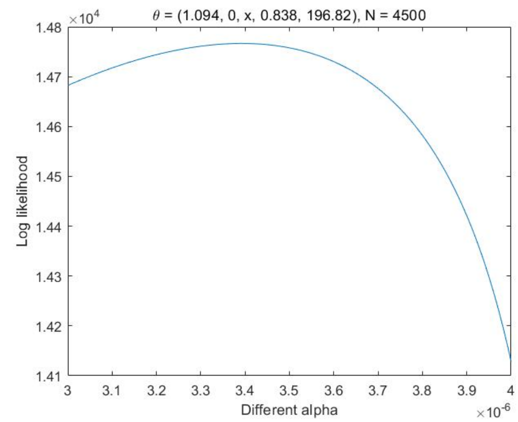
The reason behind such correlation is because estimating λ is similar to estimating interest rate r. When we have a sample average of 0.02 for the noises, then the stock prices will increase more than their expected value, and a higher λ is more likely to explain this increase, making the maximum likelihood estimate higher than the true value. Also, if we look at the Heston-Nandi model

$$R(t) = r + h(t)\lambda + \sqrt{h(t)}z(t), \tag{4.5}$$

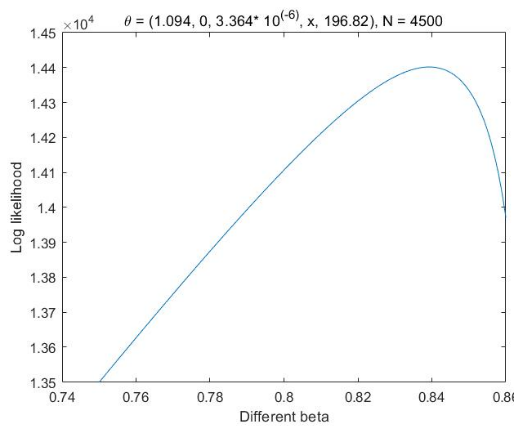
the daily variance $h(t)$ is very small, say 10^{-4}, hence the square root of $h(t)$ will be much bigger than $h(t)$ (in this case the value is 10^{-2}). Thus, the impact of λ to daily log return is lesser than the particular random noises $z(t)$ generated. In order to confirm that it is indeed the return structure (4.5) that causes this, rather than the specification of the conditional variance GARCH model, we do another experiment by setting $\alpha = \beta = \gamma = 0$, so we have constant conditional variances $h(t) = \omega$ for



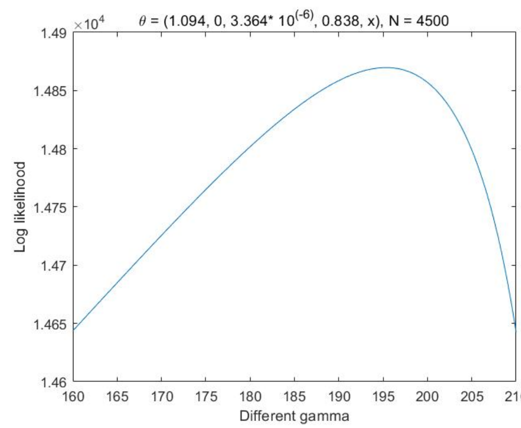
(a) Partial log likelihood w.r.t λ, $N = 4500$



(b) Partial log likelihood w.r.t α



(c) Partial log likelihood w.r.t β



(d) Partial log likelihood w.r.t γ

Figure 4.3: Partial log likelihood functions with respect to each parameter

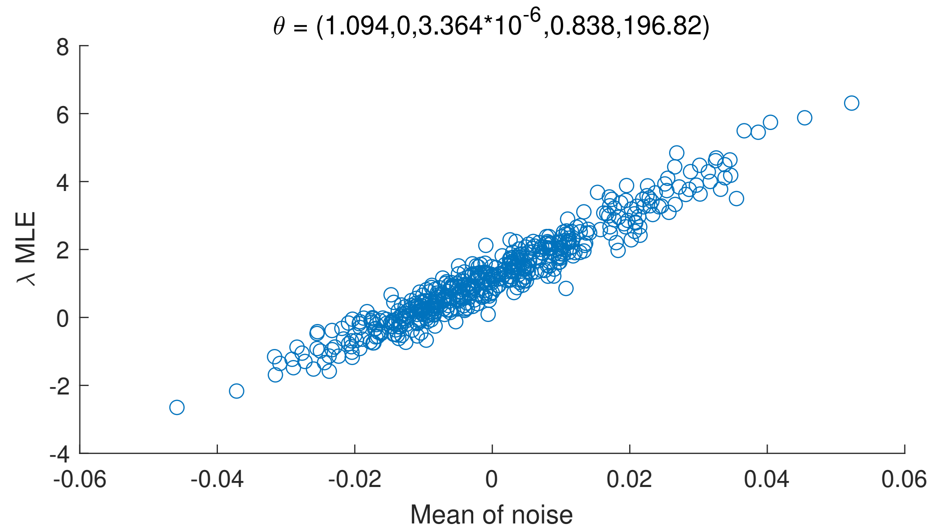


Figure 4.4: Scatter plot of mean of noise versus λ MLE, for $N = 5000$ data sample, $M = 500$ samples

all t. The following table summarizes the estimation result when the volatility is constant.

Table 4.3: Constant volatility, only estimating λ, true $\lambda = 5$, N=5,000

	Sample 1	Sample 2	Sample 3	Sample 4
True λ	5	5	5	5
Estimated λ	4.976	6.326	0.353	7.480
Mean of noise ($\times 10^{-2}$)	-0.022	1.322	-4.616	2.463

In the above experiment, we set the variance ω to be 1.062e-4, which is the long-run level of the C-H-J 2013 parameters. We simulate $N = 5,000$ data and only estimate λ (i.e. assuming we know the true volatility ω). The true λ is set to be 5, which is larger than the ordinary price of risk observed in the market. Larger true values should help the estimation, however, we still see a large variety of different samples, and a clear correlation with the mean of noise.

Note that under the assumption that the noises are i.i.d. $N(0,1)$ distributed, the mean of noise for a $N = 5000$ sample has a $N(0, 1/5000)$ distribution. A random sample of size 10 of such distribution is:

$-0.471,\ -0.094,\ 0.511,\ 0.491,\ 0.694,\ -1.342,\ -1.302,\ 1.035,\ 1.402,\ 2.752\ \ (\times 10^{-2})$

Hence, we see that the mean of noise is reasonably distributed, i.e. the mean of noise being this far from 0 should be expected, and not because of the quality of the random number generator.

Estimation of λ with ordinary least squares Next, we compare the MLE with ordinary least square (OLS) method. Luckily, when true volatility is a known constant, the OLS of λ has an analytical solution. Similarly as before, when volatility $h(t) = \omega$ is constant, the model becomes

$$R(t) = \lambda\omega + \sqrt{\omega}z(t), \tag{4.6}$$

where $R(t)$ is the daily log return, and $z(t)$ is iid standard normal noise. In order to find λ that minimize sum of squared error

$$S = \sum_{i=1}^{N} \left(\sqrt{\omega}z(i)\right)^2 = \sum_{i=1}^{N} \left(R(i) - \lambda h(i)\right)^2, \tag{4.7}$$

we have

$$\hat{\lambda} = \frac{\partial S}{\partial \lambda} = \frac{\sum_{i=1}^{N} R(i)\omega}{\sum_{i=1}^{N} \omega^2} = \frac{\overline{R(t)}}{\omega}, \tag{4.8}$$

and

$$\operatorname{Var}(\hat{\lambda}) = \frac{1}{\omega^2 N^2} \operatorname{Var}\left(\sum_{i=1}^{N} R(i)\right) = \frac{N\omega}{\omega^2 N^2} = \frac{1}{N\omega}, \tag{4.9}$$

so

$$\operatorname{SE}(\hat{\lambda}) = \frac{1}{\sqrt{N\omega}}. \tag{4.10}$$

Now, we can simulate $N = 3475$ data with $\lambda = 5$, $\omega = 10^{-4}$ (corresponds to 15% annualized standard deviation) to obtain OLS estimates:

Table 4.4: OLS of λ with true daily volatility $h(t)$

	True value	OLS	SE[1]	Mean of noise	Matlab's `fitlm`[2]			
					Estimate	SE	t-Stat	p-Value
λ	5	6.529	1.696	0.01529	6.529	1.701	3.839	0.000126

[1]Calculated based on (4.10).

[2]Output from Matlab's `fitlm` linear regressor.

Note that a standard error of 1.696 is the theoretical standard deviation of the estimator, and the result from Matlab is the sample standard error.

To summarize, we showed that λ is hard to estimate at a sample size of $N = 5000$ level. At first, we observed that the log-likelihood function is not sensitive to λ changes. Next, we observed that there is a strong correlation between the mean of noise and λ estimates. The variation of the mean of noise is the cause of the variation of λ estimates, and this is further confirmed by OLS. The variation of the mean of noise at $N = 5000$ level is too big to give a statistically meaningful estimate of λ. At a level of $N = 500000$, with a large λ, say $\lambda = 5$, the estimation becomes significant. In estimation with real data, another error source is added. Note that we assume the risk-free rate to be constant throughout the period, however, this assumption is very unrealistic. With such a vague estimate of risk-free rate we certainly cannot expect a reliable estimate of λ. Combining what we have previously, in our simulation study, we find that using a long stock return series (with around 18 years of data) can estimate α, β, γ, as well as GARCH persistence and long-run variance, at a maximum of 15% relative error, but the estimation of λ is not significant. This motivates us to bring in option data to examine the benefits of using option prices in capturing model parameters, including the risk premium λ.

4.3 Options-only calibration

Calibrating our model using option prices seems to be a natural idea, however, before we step into the joint estimation using returns and options, we explore some properties and challenges of model calibration with option prices only. The calibration with options has two immediate differences from the returns MLE. First, we cannot distinguish the parameters λ and γ only from options, because option prices only depend on the risk-neutral parameter $\gamma^* = \gamma + \lambda + 1/2$. Hence, we only have 4 model parameters to be estimated. Second, we have seen in Proposition 2.6 that parameter ω cannot take negative values. Because the ω reported from various sources (for example, $3.038e$-9 from Christoffersen, Heston and Jacobs (2006), and $2.83e$-12 from Christoffersen, Feunou, Jacobs and Meddahi (2012)) are very close to the boundary, this means the objective function only changes by a little when evaluated at, say, $\omega = 3.038e$-9 and $\omega = 0$. As a result, when we use the unconstrained optimizer `fminsearch`, the function would go into the negative ω region, causing the optimization function to abort. Because the particular focus of the simulation study is to generate many samples and obtain individual estimates, we would require the numerical optimization procedure to be robust, and we need to ensure the optimization gives global optimum, rather than local optimum. We state that the calibration time with options is usually in hours, with a single CPU. As a consequence, we cannot have 500 initial values to enhance the robustness.[2] This motivates us to follow Christoffersen, Heston and Jacobs (2013) to restrict $\omega = 0$ for the entire thesis to provide quick and stable calibrations (see Appendix C for a discussion of incorporating ω in the calibration).

4.3.1 Plot of the objective function

From the preliminary work of model calibration with option prices, we have reached many difficulties and failures with numerical optimization giving bizarre results far from the true value. Therefore, it is worth comparing to doing calibration under returns MLE, where the local optimizer almost always produce a convergent solution of local minimum, while when involving options, we find that different initial values are very likely to produce a different result, as observed as well by Kanniainen, Lin and Yang (2014). This makes us wonder about the shape of the objective function. During one of the experiments of calibrating $N = 1200$ simulated options with an initial value of

$$(\lambda, \omega, \alpha, \beta, \gamma) = (-, 0, 3e\text{-}6, 0.05, 200), \tag{4.11}$$

we obtain the following parameter estimation, as seen in Table 4.5, Estimate 1. From the table, we see that even when the parameters are far off from the true values, we do get fairly close persistence, and long-run expected daily variance. This provides the first hint that the objective function (vega-weighted RMSE) has many local minimums, and a necessary condition is that the persistence and long-run expected variance are close to the true ones. This conjecture is valid because options only capture the average (or expected) movement of spot asset price within a period, whereas from spot prices we can filter out a conditional volatility series and hence extract each GARCH parameter.

[2] Kanniainen, Lin and Yang (2014) calibrated one option sample with 500 randomly selected initial values to remedy the defects for using a local optimizer. It would be computationally cumbersome to have 500 initial values for 100 samples.

Table 4.5: Vega-weighted least squares, $N = 1200$, starting at $(-, 0, 3e\text{-}6, 0.05, 200)$

	True value	Estimate 1	Estimate 2 (Corrected)
λ	2.772	-	-
ω	3.038e-9	3.007e-6	4.108e-7
α	3.66e-6	6.047e-7	3.252e-6
β	0.9026	0.6706	0.9
γ	128.4	659.678	139.069
$\beta + \alpha\gamma^2$	0.9629	0.9632	0.9632
$\frac{\omega+\alpha}{1-\beta-\alpha\gamma^2}$	9.884e-5	9.819e-5	9.819e-5

To validate the conjecture, we make a surface of our objective function. Since we have 4 risk-neutral parameters in HN-GARCH, we can make a surface where the x and y axis are different values of β and γ, and we imply ω and α by the following relations

$$\begin{cases} \beta + \alpha\gamma^2 = 0.9632, \\ \frac{\omega+\alpha}{1-\beta-\alpha\gamma^2} = 9.819 \times 10^{-5}, \end{cases} \tag{4.12}$$

and evaluate the objective function based on the parameters. The objective function surface (Figure 4.5) clearly shows an increasing line where the objective function is very low, however, when γ increases from 100 to 600, the lowest point still increases, as we see the slope is not parallel to the $xy-$plane. As a result, the global minimum is at the region when γ is between 100 and 200, which is where the true parameter lies. We note that the parameter sets on this surface all satisfy the conditions (4.12), hence, we conclude that the full objective function (of 4 model parameters) indeed has many local minimums where the optimization algorithm could easily get stuck with.

Repeating the same exercise with the C-H-J 2013 parameters shows a comparable results. In Figure 4.6, we plot the surface of vega-weighted Gaussian log-likelihood versus different value of β and γ, and imply ω and α by the following relations

$$\begin{cases} \beta + \alpha\gamma^2 = 0.9683, \\ \frac{\omega+\alpha}{1-\beta-\alpha\gamma^2} = 1.0617 \times 10^{-4}, \end{cases} \tag{4.13}$$

where these values are the persistence and long-run daily conditional variance completed at true parameter value (the parameters in 2.18). Compared to Figure 4.5, the difference is that now the z-axis shows the (negative of) log-likelihood values of the option sample, and the z-axis is capped at -3600, for demonstration purposes. By comparing the front and back view in Figure 4.6, the function shape is similar to 4.5. Furthermore, from angle 3, we find two points which correspond to pairs of $(\beta, \gamma) = (0.84, 200)$ and $(\beta, \gamma) = (0.66, 580)$ whose parameters differ greatly but the log-likelihood only shows a 3 difference in absolute value. This also confirms with our previous findings that the non-reduced objective function (with 4 HN-GARCH parameters as inputs) must contain a great number of local minimums, making it difficult to optimize the function numerically.

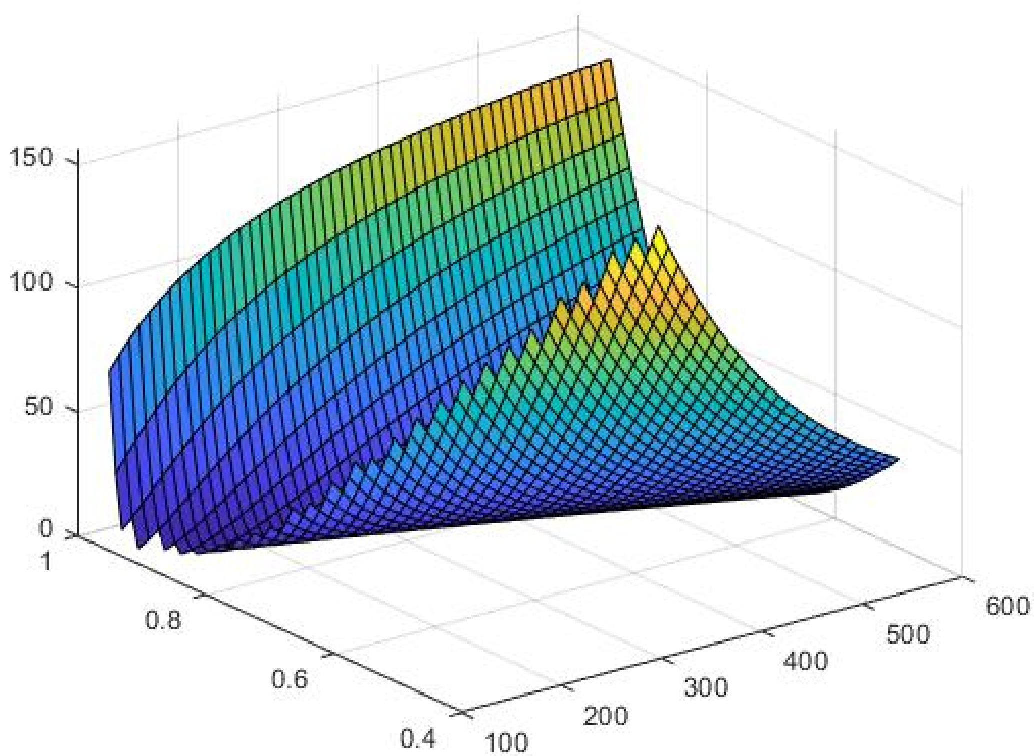
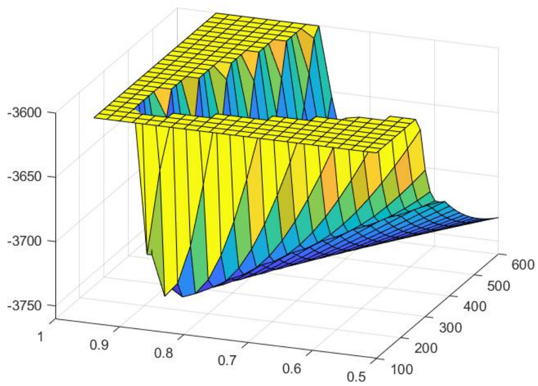
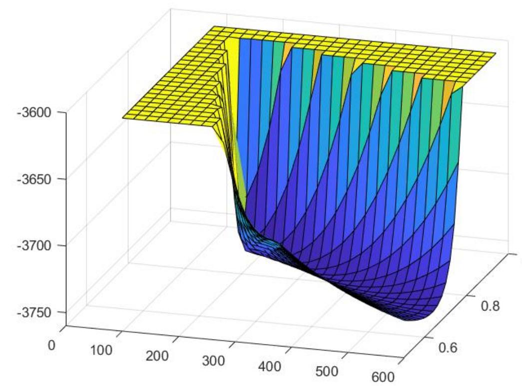


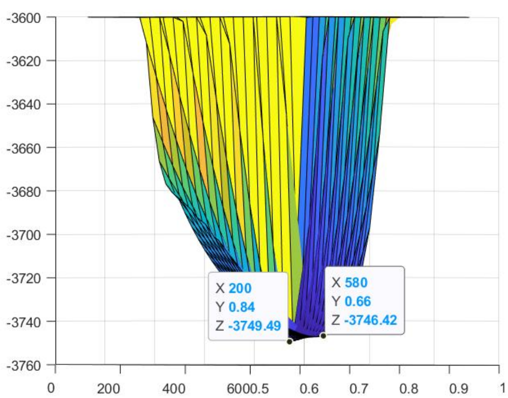
Figure 4.5: Surface of vega-weighted RMSE versus β and γ, with fixed persistence and long-run expected variance



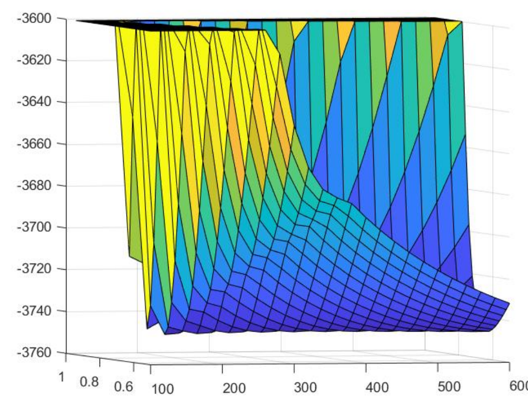
(a) Angle 1 (front)



(b) Angle 2 (back)



(c) Angle 3 (back and horizontal)



(d) Angle 4 (side)

Figure 4.6: Surface of vega-weighted Gaussian log-likelihood versus β and γ, with fixed persistence and long-run expected variance

4.3.2 Numerical fixes

Motivated by the above observations, when the numerical optimization encounters a local minimum, we can always use a three-step procedure to correct the optimization result. The procedure is summarized as follows:

Step 1. We first estimate the persistence, $\beta + \alpha\gamma^2$. This is achieved by replacing β by $\beta + \alpha\gamma^2$ in the objective function. Theoretically, this should not change the optimization, however, we can manually make the algorithm focus more on estimating persistence by increasing its scale in the parameter scaling process.

Step 2. Next, we estimate long-run expected volatility with fixed persistence. This is done by replacing one parameter with the long-run volatility (usually α).

Step 3. Last, we estimate the remaining 2 parameters with fixed persistence and long-run expected volatility to get better estimation results.

In brief, the three-step procedure takes into account the special structure of the objective function (as seen in Figure 4.5) and provides a numerical trick to overcome the local minimum issue. Note that this procedure can be applied with or without setting $\omega = 0$. We present the calibration result using the three-step procedure in Table 4.5 (Estimate 2). By comparing the two estimation results we see the benefits of applying the three-step procedure.

Although the three-step procedure brings better results, it is numerically complicated to carry out three rounds of optimizations. Instead, considering the shape of the objective function, we can apply a numerical transformation that instead estimates the maximum likelihood persistence, long-run variance instead of β and α respectively. This replacement is applied because β accounts for most of the portion in persistence, while α dominates ω in the long-run variance formula. The transformation invariance property of MLE implies that when obtaining the maximum likelihood values for persistence and long-run variance, we can transform back to obtain the MLE for β and γ. To ensure we truly get the global optimum from the numerical optimizer, for all the optimizations involving options, we use a fixed grid approach to set 3 different initial values for λ, γ and long-run variance, giving 27 initial values for each sample. From our result, most of the initial values, if not all, terminated to the same value after the optimization algorithm, which confirms the effectiveness of doing the one-to-one transformation of parameter mapping, and further validates our results.

Comparison of optimization result

Table 4.6 demonstrates the calibration result under the "Options-only calibration, method 1" approach, with the Gaussian log likelihood of vega weighted errors as the objective function, and the true daily conditional variance $h(t)$ are assumed to be known and directly used. Hence, we do not estimate or update the variance during the optimization process. In comparison, Table 4.7 includes the result when we do filter daily conditional variances $h(t)$ from the return series. The option sample are simulated based on the true parameters (C-H-J 2013 parameters, see (2.18) and (2.19)) and the following specification (see Table 4.1 for the option sample statistics):

$$N = 500 \text{ specifications:} \begin{cases} K = 95, 100, ..., 115, \\ T = 23, 46, \\ h(5), h(10), ..., h(250), \end{cases} \tag{4.14}$$

where the $h(t)$ are the daily conditional variances from the simulated return series, and the 5-day lag is understood as if the option prices are observed on each Wednesday. The true option prices are imposed with vega errors from an i.i.d. $N(0, 4.95\%)$ distribution, and the $N = 1500$ and $N = 4500$ sample used the same specification but generate the above option set for 3 and 9 times with different random error.[3]

From both tables, we can indeed see that, as we increase the number of options, the estimates are closer to the true values, which validates the consistency. The result with $N = 4500$ options is fairly close to the true values. Compared with Returns-only MLE results, we see that using $N = 1500$ options has a comparable sample standard deviation to using $N = 4500$ returns, which suggests that each single option observation provides more information and predictability than one piece of return. From both tables, the small sample ($N = 500$) estimates tend to show a 10% relative error to the true value, however, the maximum likelihood long-run variances only show 9.6% and 1.65% relative difference, respectively in Table 4.6 and Table 4.7. Moreover, the relative difference in persistence is at 0.63%. Hence, we conclude that option prices are more sensitive to persistence and long-run expected variance, rather than single parameters.

When comparing method 1 with method 2, we see that the sample standard error for γ is greater in method 1, while the sample standard error for α and β are lower in method 1 for all sample sizes (except for α with $N = 500$, where the difference is minimal). Because this pattern is across all sample sizes, we might conclude that γ

[3]The number 4.95% is the vega weighted RMSE in Kanniainen, Lin and Yang (2014).

Table 4.6: Estimation result: Options-only calibration method 1 (known $h(t)$)

Parameter	True value	$N = 500$		$N = 1500$		$N = 4500$	
		MLE (SE)	Sample SE	MLE (SE)	Sample SE	MLE (SE)	Sample SE
ω	0	0 -	-	0 -	-	0 -	-
α	3.364e-6	3.671e-6 (6.063e-7)	6.620e-7	3.672e-6 (3.405e-7)	2.824e-7	3.537e-6 (2.109e-7)	2.007e-7
β	0.838	0.818 (2.200e-2)	2.216e-2	0.841 (1.200e-2)	1.239e-2	0.838 (0.737e-2)	0.660e-2
γ	196.82	204.24 (23.151)	38.380	182.14 (8.524)	16.366	190.10 (6.172)	7.924
$\beta + \alpha\gamma^2$	0.968	0.971		0.963		0.965	
Annualized Vol.	16.36%	17.93%		15.87%		16.04%	

The above table shows the option calibration results with different sets of options using the Gaussian log-likelihood function of the vega weighted errors as the objective function. The option data sets are generated based on the true parameters (C-H-J 2013 parameters, see (2.18) and (2.19)) and with the specification in (4.10). The calibration is conducted with ω restricted to 0, to enhance the robustness of the numerical optimization (this is also applied by Christoffersen, Heston and Jacobs (2013)). Note that the risk-neutral parameter $\gamma^* = \gamma + 1.094 + 0.5$ is used in finding option prices, where 1.094 is the true λ. Since option prices cannot identify λ, this number is kept fixed throughout the calibration. The estimated standard errors are computed via the outer product of the gradient and are shown in parenthesis. The sample standard error is computed by sample standard deviation of the optimized parameters for 20 simulated samples for each N. The persistence ($\beta + \alpha\gamma^2$) and annualized volatility (computed via (2.10)) at MLE parameters are also included.

Table 4.7: Estimation result: Options-only calibration method 2 ($h(t)$ filtered from return series)

Parameter	True value	$N = 500$		$N = 1500$		$N = 4500$	
		MLE (SE)	Sample SE	MLE (SE)	Sample SE	MLE (SE)	Sample SE
ω	0	0 -	-	0 -	-	0 -	-
α	3.364e-6	4.025e-6 (6.764e-7)	6.654e-7	3.311e-6 (3.318e-7)	3.881e-7	3.356e-6 (1.987e-7)	2.269e-7
β	0.838	0.828 (2.561e-2)	2.222e-2	0.855 (1.184e-2)	1.196e-2	0.845 (0.745e-2)	0.696e-2
γ	196.82	181.99 (12.498)	24.116	184.53 (8.623)	12.775	192.47 (5.813)	7.385
$\beta + \alpha\gamma^2$	0.968	0.961		0.967		0.968	
Annualized Vol.	16.36%	16.09%		15.88%		16.11%	

The above table shows the option calibration results with different sets of options using the Gaussian log-likelihood function of the vega weighted errors as the objective function, and the daily variances $h(t)$ are filtered from return observations. The option data sets are generated based on the true parameters (C-H-J 2013 parameters, see (2.18) and (2.19)) and with the specification in (4.10). The calibration is conducted with ω restricted to 0, to enhance the robustness of the numerical optimization (this is also applied by Christoffersen, Heston and Jacobs (2013)). Note that the risk-neutral parameter $\gamma^* = \gamma + 1.094 + 0.5$ is used in finding option prices, where 1.094 is the true λ. Since option prices cannot identify λ, this number is kept fixed throughout the calibration. The estimated standard errors are computed via the outer product of the gradient and are shown in parenthesis. The sample standard error is computed by sample standard deviation of the optimized parameters for 20 simulated samples for each N. The persistence ($\beta + \alpha\gamma^2$) and annualized volatility (computed via (2.10)) at MLE parameters are also included.

benefits from the volatility updating and filtering, which provides additional information (or sensitivity to γ), resulting in the lower sample standard error, whereas, α and β benefit from having the true volatility.

In terms of the finite sample behaviour of standard error estimates, we see sample standard deviations are in general reasonably close to the estimated standard errors, computed at the maximum likelihood value, showing the validity of the formula. For small samples, since the maximum likelihood estimates deviate a lot from the true value, the standard error estimates are not so accurate. As we increase the sample size, however, the standard error estimates are very close to sample standard deviations, both relatively and absolutely.

4.4 Joint estimation-calibration with returns and options

Now we combine the return and option observations to demonstrate the effectiveness of the joint estimation. The results are shown in Table 4.8. In order to test the asymptotic properties, we triple the number of returns and options for each sample size, without changing the ratio between the two, to establish a fair comparison. From the table, we see that the estimation results are getting more accurate when N is increased, which confirms consistency, Also, the estimated standard errors (from the outer product of gradients) are close to the sample standard deviations of the estimators. The relative difference of the estimated standard errors between the sample realizations is comparable as in the return and option-only cases, hence confirming the validity of such estimation approach. This result is expected since it is consistent with both returns- and options-only cases.

Because the return observation and option observations carry different information regarding each parameter, one natural study for the joint estimation is to compare the joint estimations with different numbers of returns and options. To answer that, we propose a study that shows the joint estimation with N_r varies from $500, 1500$, and 4500 observations, and N_o varies from $0, 500, 1500$, and 4500 observations, for a total of $3 \times 4 = 12$ blocks. When $N_o = 0$, the joint estimation will be reduced to Returns-only estimation. For each block, we simulate the returns and options using the same seed (the random numbers generated for smaller samples will be present again in larger samples) for 100 samples, and report the RMSE for each parameter as in Table 4.10. We do the same simulation, but only changing the standard deviation of vega-weighted noise to 1.24 % (4 times smaller than in Table 4.10) and 19.84% (4 times larger than in Table 4.10). In this way, when we compare both tables, we also see how the joint estimation behaves when observed option prices are more noisy.

Table 4.8: Estimation result: Joint estimation

Parameter	True value	$N_r = N_o = 500$		$N_r = N_o = 1500$		$N_r = N_o = 4500$	
		MLE (SE)	Sample SE	MLE (SE)	Sample SE	MLE (SE)	Sample SE
λ	1.094	3.1671 (5.3151)	4.381	3.843 (2.784)	2.351	1.211 (1.469)	1.338
ω	0	0 -	-	0 -	-	0 -	-
α	3.364e-6	3.531e-6 (5.001e-7)	5.460e-7	3.240e-6 (2.637e-7)	3.248e-7	3.292e-6 (1.554e-7)	1.741e-7
β	0.838	0.850 (1.742e-2)	1.872e-2	0.854 (0.981e-2)	1.008e-2	0.840 (0.638e-2)	0.542e-3
γ	192.84	178.11 (13.890)	22.975	184.47 (9.179)	11.282	197.27 (5.941)	6.280
$\beta + \alpha\gamma^2$	0.968	0.962		0.965		0.969	
Annualized Vol.	16.38%	15.27%		15.20%		16.14%	

The above table shows the Joint estimation-calibration results with different numbers of returns and options using the Gaussian joint likelihood (3.23) as the objective function. The N_r and N_o are the number of returns and options in the sample, respectively. We increase the number of returns and options without changing their ratio in the sample. The option data sets are generated based on the true parameters (C-H-J 2013 parameters, see (2.18) and (2.18)) and with the specification in (4.8). The calibration is conducted with ω restricted to 0, to enhance the robustness of the numerical optimization (this is also applied by Christoffersen, Heston and Jacobs (2013)). Note that the implied risk-neutral parameter $\gamma^* = \gamma + \lambda + 0.5$ is used in finding option prices. The estimated standard errors are computed via the outer product of the gradient and are shown in parenthesis. The sample standard error is computed by sample standard deviation of the optimized parameters for 100 simulated samples for each N. The persistence $(\beta + \alpha\gamma^2)$ and annualized volatility (computed via (2.10)) are computed at MLE values.

Table 4.9: Comparison of the maximum likelihood estimator's sample RMSE (of 100 samples) for different number of returns and options in joint estimation. Data is simulated based on C-H-J 2013 parameters. Identical random seeds are used (the seed for small samples are nested in big samples). The standard deviation of vega noise is 1.24%.

Sample RMSE	No. of options	0	500	1500	4500
No. of returns					
500	λ	4.4204	4.3849	4.3721	4.3716
	α	1.1512e-06	1.6084e-07	9.4525e-08	5.7253e-08
	β	4.3444e-02	5.2676e-03	2.9494e-03	1.7749e-03
	γ	58.3669	6.9269	4.8843	4.5805
1500	λ	2.3516	2.3507	2.3495	2.3481
	α	6.4991e-07	1.5809e-07	9.5280e-08	5.7460e-08
	β	2.3160e-02	5.3681e-03	2.8942e-03	1.7515e-03
	γ	27.9832	5.6402	3.7619	2.8203
4500	λ	1.3329	1.3310	1.3314	1.3328
	α	3.2247e-07	1.4891e-07	9.3197e-08	5.6282e-08
	β	1.3335e-02	5.0599e-03	2.8306e-03	1.7078e-03
	γ	16.4485	4.9961	3.3570	2.2458

Table 4.10: Comparison of the maximum likelihood estimator's sample RMSE (of 100 samples) for different number of returns and options in joint estimation. Data is simulated based on C-H-J 2013 parameters. Identical random seeds are used (the seed for small samples are nested in big samples). The standard deviation of vega noise is 4.96%.

Sample RMSE	No. of options	0	500	1500	4500
No. of returns					
500	λ	4.4204	4.3754	4.3692	4.3979
	α	1.1512e-06	5.4438e-07	3.6385e-07	2.1718e-07
	β	4.3444e-02	1.8963e-02	1.1673e-02	6.8525e-03
	γ	58.3669	23.0346	12.5457	8.1215
1500	λ	2.3516	2.3480	2.3450	2.3523
	α	6.4991e-07	4.3240e-07	3.2572e-07	2.0948e-07
	β	2.3160e-02	1.6352e-02	1.0387e-02	6.3956e-03
	γ	27.9832	16.1635	11.2781	7.1267
4500	λ	1.3329	1.3308	1.3307	1.3345
	α	3.2247e-07	2.9597e-07	2.4395e-07	1.7329e-07
	β	1.3335e-02	1.0996e-02	8.1642e-03	5.4059e-03
	γ	16.4485	12.1714	9.2328	6.2579

Table 4.11: Comparison of the maximum likelihood estimator's sample RMSE (of 100 samples) for different number of returns and options in joint estimation. Data is simulated based on C-H-J 2013 parameters. Identical random seeds are used (the seed for small samples are nested in big samples). The standard deviation of vega noise is 19.84%.

Sample RMSE	No. of options	0	500	1500	4500
No. of returns					
500	λ	4.4204	4.3751	4.3448	4.4203
	α	1.1512e-06	9.2688e-07	8.5886e-07	6.5108e-07
	β	4.3444e-02	3.9436e-02	3.4136e-02	2.4259e-02
	γ	58.3669	51.5553	41.9048	28.9388
1500	λ	2.3516	2.3473	2.3412	2.3510
	α	6.4991e-07	6.3342e-07	5.9599e-07	4.9468e-07
	β	2.3160e-02	2.2719e-02	1.9914e-02	1.6059e-02
	γ	27.9832	26.0200	23.1370	19.0865
4500	λ	1.3329	1.3319	1.3317	1.3357
	α	3.2247e-07	3.3707e-07	3.3724e-07	2.9666e-07
	β	1.3335e-02	1.2950e-02	1.2262e-02	1.0561e-02
	γ	16.4485	16.0470	15.2992	13.4823

From the blocks in both tables above, if we compare each column or row, we see that when the number of options is fixed, increasing returns will lead to lower RMSEs. The same will apply when we fix the number of returns as well. This should be expected, because as we bring in more data, we expect to have estimates with lower error, hence lower RMSEs. This observation is consistent across all sample sizes, and for parameters α, β, and γ. For λ, however, whenever we increase the number of returns, we see a decrease in RMSE. When compared horizontally, the RMSE of λ stays at a similar level. From this, we conclude that there is no evidence suggesting that adding only options can help capture the risk premium parameter λ. This result implies that option sample, although further provides information on $\lambda + \gamma$, cannot help identify λ.

If we compare the diagonal symmetrical blocks in Table 4.10, we see that when the total number of observations fixed, the samples with more options have lower RMSE for all the GARCH parameters, except for λ. This implies that, regarding α, β and γ, the return observations are noisier than option prices with the option vega noise set to 4.96%, and we are better off using options to capture the model parameters. The same conclusion can be drawn for Table 4.9, since the option noise is even smaller. As a comparison, in Table 4.9, once we increase the option noise level, we see that the estimation favors return data. With the default noise level as in Table 4.11, when adding an option sample that is a third of the existing return sample, for example, 1500 returns plus 500 options, or 4500 returns plus 1500 options, the RMSEs of GARCH parameters are halved, indicating the importance and efficiency of adding an option sample. Even with quadrupled noises in Table 4.11, using 1500 returns plus 4500 options can still achieve comparably small RMSEs as using 4500 return series. Generally, since GARCH parameters could change over years, we could get misleading information when fitting a long return series. The results in Table 4.11 demonstrates that a short return series plus a large option sample can achieve sufficient accuracy, hence should be suggested over fitting a long return series.

4.4.1 Summary

In summary, we have shown empirically the consistency of the maximum likelihood estimators for model parameters and the outer product of gradient for their standard errors for the three types of estimations. Although a joint estimation with returns and options does not help in capturing the risk premium factor, λ, the benefits of adding an option sample into the joint estimation is noticeable to parameters α, β, and γ. In particular, when we compare the Returns-only MLE with the joint estimation, we see that adding option observations to return observations will help the estimation significantly when the option sample is less noisy (4.96% standard deviation on the vega-weighted option noise). When we compare the calibration result when using only $N_r = 4500$ returns, and $N_r = 4500$ returns with additional $N_o = 4500$ options, we see that the sample standard errors decreases by 4 times for parameters α, β, and γ (the standard errors for sample implied persistence and annualized volatility also decreased by the same scale). On the other hand, once we have more noisy option data, the estimation favors return observations, but the estimation still benefits from having the additional option sample. This verifies that one should fit GARCH option pricing models jointly with a short return series plus a large option data.

Chapter 5

Conclusion

In this thesis, we explored open problems regarding the Heston-Nandi model and its estimation procedures. In the existing literature, we have seen many works, for example, Heston, and Nandi (2000), and Christoffersen, Heston, and Jacobs (2013) that incorporated real option prices into model calibration with various approaches. However, there have been no studies on the empirical behavior and theoretical benefits of calibrating the HN-GARCH model with option prices, in a controlled experiment. Motivated by this, we did an extensive amount of simulation-based studies and analysis, and confirmed the validity of the joint estimation approach. In particular, the results actively demonstrate that, by adding the same number of options in the existing return sample, we could lower the standard error of the MLE estimates of α, β, and γ by 2 to 4 times, depending on the nosiness of the option data. As a consequence, the implied persistence and annualized volatility become more accurate by the same scale. Furthermore, we demonstrated empirically the consistency of the joint estimation and the standard error formula. In addition, we compared the information that return and option data carries with respect to each parameter, and illustrated the consequences when we add noisy data to the estimation. As one expects, when the option sample is more noisy, the Joint estimation-calibration gains more accuracy with additional return data over option data. However, even with a noisy option data set, the joint estimation still benefits greatly from the additional option observations. This suggests that one should fit GARCH option pricing models with a medium-sized return series (of 6-8 years) and a large option sample, instead of fitting the model with a very long return series. On the downside, since λ is not identifiable with additional option data, the joint estimation with options cannot capture more information of λ. As a consequence, with the assumed HN-GARCH model, we can only reply on returns to estimate the price of risk.

Besides the concrete end-goal, we have reached many meaningful conclusions during the discovery process. First, we give proofs of why the GARCH parameters, especially ω, cannot take negative values. Second, we presented numerical tricks and details, outlined several numerical challenges and their potential solutions. Lastly, although this work is in progress due to time constraints, we studied which type of options is the most effective in capturing HN-GARCH parameters, and concluded that short-maturity, out-of-money options are the most valuable assets in an option sample (see Appendix A).

Based on this research, we have several interesting problems for future directions. These directions lie in many categories. For example:

1. (Numerical optimization) Is there a robust, computationally feasible way to estimate the model with parameter ω (that does not rely on multiple initial values)? How robust are the numerical tricks outlined in Appendix C?

2. (Model structure) The "vega" assumption essentially assumes the implied volatility errors are identical and independently distributed. How realistic is this assumption? Will the result stays the same if we change the option noise structure from vega noise to a relative noise?

3. (Model structure) We cannot distinguish λ and γ from option prices. Hence, options do not provide much value in estimating λ. Is it realistic that risk premium does not appear in option prices? For other GARCH-type pricing models where λ is identifiable in option prices, how well can option data capture lambda?

4. (Estimation) Which options, in terms of K, T and daily conditional variance $h(t)$, are the most effective in model calibration?

We hope we can answer all these questions in future works.

References

[1] Jinho Bae, Chang-Jin Kim, and Charles R Nelson. Why are stock returns and volatility negatively correlated? *Journal of Empirical Finance*, 14(1):41–58, 2007.

[2] Geert Bekaert and Guojun Wu. Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, 13(1):1–42, 2000.

[3] Ernst R Berndt, Bronwyn H Hall, Robert E Hall, and Jerry A Hausman. Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4*, pages 653–665. NBER, 1974.

[4] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[6] Mark Broadie, Mikhail Chernov, and Michael Johannes. Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, 62(3):1453–1490, 2007.

[7] John Y Campbell and Ludger Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318, 1992.

[8] R Cavazos-Cadena and G González-Farıas. Necessary and sufficient conditions for the consistency of maximum likelihood estimators. *Advances and Applications in Statistical Sciences*, 3(1):13–42, 2010.

[9] Stanley H. Chan. Ece 645: Estimation theorey lecture 8: Properties of maximum likelihood estimation (mle). https://engineering.purdue.edu/ChanGroup/ECE645Notes/StudentLecture08.pdf, 2015. Accessed: 2020–12-24.

[10] Christophe Chorro and Rahantamialisoa H Fanirisoa. Option valuation with ig-garch model and a u-shaped pricing kernel. *Soft Computing*, 24(12):8505–8522, 2020.

[11] Andrew A Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432, 1982.

[12] Peter Christoffersen, Bruno Feunou, Kris Jacobs, and Nour Meddahi. The economic value of realized volatility: Using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis*, pages 663–697, 2014.

[13] Peter Christoffersen, Steve Heston, and Kris Jacobs. Option valuation with conditional skewness. *Journal of Econometrics*, 131(1-2):253–284, 2006.

[14] Peter Christoffersen, Steven Heston, and Kris Jacobs. Capturing option anomalies with a variance-dependent pricing kernel. *The Review of Financial Studies*, 26(8):1963–2006, 2013.

[15] Peter Christoffersen and Kris Jacobs. The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291–318, 2004.

[16] Peter Christoffersen and Kris Jacobs. Which garch model for option valuation? *Management Science*, 50(9):1204–1221, 2004.

[17] Harald Cramér. *Mathematical methods of statistics*, volume 43. Princeton university press, 1999.

[18] Martin J Crowder. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(1):45–53, 1976.

[19] Jin-Chuan Duan. The garch option pricing model. *Mathematical Finance*, 5(1):13–32, 1995.

[20] Bradley Efron and David V Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–483, 1978.

[21] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

[22] Marcos Escobar, Javad Rastegari, and Lars Stentoft. Affine multivariate garch models. Technical report, Working Paper, 2019.

[23] Jinji Hao and Jin E Zhang. Garch option pricing models, the cboe vix, and variance risk premium. *Journal of Financial Econometrics*, 11(3):556–580, 2013.

[24] Ludger Hentschel. Errors in implied volatility estimation. *Journal of Financial and Quantitative Analysis*, pages 779–810, 2003.

[25] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[26] Steven L Heston and Saikat Nandi. A closed-form garch option valuation model. *The Review of Financial Studies*, 13(3):585–625, 2000.

[27] Bruce Hoadley. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics*, pages 1977–1991, 1971.

[28] Dandar (https://stats.stackexchange.com/users/16853/dandar). Observed information matrix is a consistent estimator of the expected information matrix? Cross Validated. https://stats.stackexchange.com/q/65597. Accessed: 2020–12-20.

[29] Eric Jacquier and Robert Jarrow. Bayesian analysis of contingent claim model error. *Journal of Econometrics*, 94(1-2):145–180, 2000.

[30] Juho Kanniainen, Binghuan Lin, and Hanxue Yang. Estimating and using garch models with vix data for option valuation. *Journal of Banking & Finance*, 43:200–211, 2014.

[31] M. Nicholas Kiefer. A review on maximum likelihood estimation. https://courses.cit.cornell.edu/econ620/reviewm5.pdf, 2008. Accessed: 2020–12-20.

[32] Sang-Won Lee and Bruce E Hansen. Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52, 1994.

[33] George Maldonado and Sander Greenland. A comparison of the performance of model-based confidence intervals when the correct model form is unknown: coverage of asymptotic means. *Epidemiology*, pages 171–182, 1994.

[34] KW Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, pages 2112–2245, 1994.

[35] Alecos Papadopoulos. The information matrix equality: proof, misspecification, and the quasi-maximum likelihood estimator. https://alecospapadopoulos.files.wordpress.com/2014/05/the-information-matrix-equality1.pdf, 2014. Accessed: 2020–12-20.

[36] Eric Renault. Econometric models of option pricing errors. *Econometric Society Monographs*, 28:223–278, 1997.

[37] Andrew A Weiss. Asymptotic theory for arch models: estimation and testing. *Econometric Theory*, pages 107–131, 1986.

[38] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

Appendix A

Which Options to Use in Option Calibration

Option prices with different strike prices, maturities and spot variances distribute differently. Therefore, various types of options should have nonparallel predictability towards the model parameters. For example, the out-of-money call options are more influenced by the tail behavior of stock returns, compared to at-the-money and in-the-money calls. In addition, spot variances play a bigger role in options with a shorter maturity, whereas for options with long maturities, the long-run level of variance is more relevant. Hence, it is natural to investigate which type of options is better in model calibration by restricting the option sample to include the specific type of options and compare the RMSEs for 100 samples.

In this experiment, we choose an intermediate level of $N = 1500$ options. As seen in (4.4), the benchmark $N = 1500$ option sample contains 500 different combinations of K, T and $h(t)$:

$$\begin{cases} K = 95, 100, 105, 110, 115, \\ T = 23, 46 \\ h(5), h(10), ..., h(250), \end{cases} \tag{A.1}$$

where we simulate 3 random errors for each combination of options. We first compare the influence by the strike prices, and the result is shown in Table A.1. To carry out the comparison, the option sample in each column all have the same strike prices, as specified in the column title, and the structure for T and $h(t)$ is unchanged as in (A.1). We generate 100 such samples, report the RMSEs of α, β, and γ, and further compare the option sample with strike prices with the benchmark sample.[1] In Table A.1, the first half refers to the RMSEs when Options-only calibration method 1 (i.e. true $h(t)$ is known and kept fixed), whereas the second half uses Options-only calibration method 2 (i.e. $h(t)$ is filtered from returns).

[1] Note that $\lambda = 1.094$ and $\omega = 0$ are kept fixed in the optimization, so the RMSEs are 0.

Table A.1: RMSEs for option samples with different strike prices

	$K = 90$	$K = 95$	$K = 100$	$K = 105$	$K = 110$	Benchmark
Method 1						
λ	0	0	0	0	0	0
ω	0	0	0	0	0	0
α	7.8594e-07	6.5765e-07	7.8206e-07	5.0149e-07	4.1100e-07	3.8133e-07
β	0.0900	0.1043	0.4242	0.0357	0.0202	0.0125
γ	96.9049	92.7515	199.5387	24.2059	14.2631	16.6518
Method 2						
λ	0	0	0	0	0	0
ω	0	0	0	0	0	0
α	4.5102e-07	4.5813e-07	4.5599e-07	4.3479e-07	5.2424e-07	3.8718e-07
β	0.0316	0.0282	0.0224	0.0169	0.0236	0.0122
γ	30.6597	25.9927	20.3174	14.9570	12.3702	12.8472

Table A.2: Calibration result

	True parameter	Sample 1 calibration result
λ	1.094	1.094
ω	0	0
α	3.364e-6	4.818e-6
β	0.838	0.235
γ	196.82	388.68
$\beta + \alpha\gamma^2$	0.968	0.962
Annualized Vol.	16.35%	17.68%

In the table, one immediately observes large RMSEs for in- and at-the-money options. In particular, for $K = 100$, the RMSE of α, β and γ is more than 2, 30 and 10 times the benchmark value. When examining the calibration results for the first sample of $K = 100$, as reported in Table A.2, we see that the implied persistence is very close to the true values, and the annualized volatility is within $\pm 1.5\%$ range of the true value, however, in terms of individual parameters, the calibrated values show large errors compared to their true values. In fact, the calibration result is also obtained using 9 different initial values, and the fact that all 9 optimizations return almost the same calibration values partly validates our result.[2] The sample 1 calibration result confirms that option prices are more sensitive to persistence and long-run variance, instead of individual parameters. When directly comparing method 1 with method 2, we see that the RMSEs for method 2 are at a comparable level with the benchmark. This is because when filtering variances from returns, a particularly small β, say 0.235 in our case, is easily rejected because the resulting (filtered) variances will not match the true variances. This shows the benefits and robustness of using Options-only calibration method 2 over method 1. Also, for method 1, the problem of large RMSEs is not shown for out-of-money samples. One probable cause is that out-of-money options capture tail behaviors (i.e. higher moments of stock returns) that might pose extra conditions on individual parameters. When comparing horizontally for Options-only calibration method 2, we see that the RMSEs for out-of-money option samples are substantially better than at-the-money and in-the-money samples. In particular, the RMSEs for α stays at a similar level across all maturities, whereas the RMSEs for β and γ reduces by 50% and 60%, respectively. Although for method 2, we see that the RMSEs for the benchmark is still higher than almost all the other samples (except γ for $K = 110$), we can still conclude that out-of-money options are beneficial in capturing β and γ, and are essential assets to be included in the option sample set, when calibrating HN-GARCH models using options.

[2] The optimizations with different initial values all converge to the same number within a 0.1% relative difference. This is consistent for all the samples.

Table A.3: RMSEs for option samples with different maturities

	$T = 23$ (1m)	$T = 69$ (3m)	$T = 138$ (6m)	Benchmark
Method 1				
λ	0	0	0	0
ω	0	0	0	0
α	4.0075e-07	3.8638e-07	6.3700e-07	3.8133e-07
β	0.0115	0.0165	0.0329	0.0125
γ	19.7730	14.8046	20.4487	16.6518
Method 2				
λ	0	0	0	0
ω	0	0	0	0
α	3.5933e-07	5.1038e-07	9.6436e-07	3.8718e-07
β	0.0107	0.0191	0.0452	0.0122
γ	13.7483	13.7895	18.4350	12.8472

Next, we repeat the experiment with different maturities. Note that the benchmark sample only includes options expiring in 1-month and 2-month. For our comparison, we include 3 additional option samples with $T = 23$ (1-month), $T = 69$ (3-month) and $T = 138$ (6-month), and the RMSEs for 100 samples are reported in Table A.3. The results for Options-only calibration method 1 show a non-monotonic pattern. In particular, for α and γ, the RMSE first decreased when we swap from 1-month options to 3-month, and then increased by 65% and 38% respectively, for 6-month options. In contrast, for β, the RMSE when swapping from 1-month to 3-month slightly increases, and then nearly doubled when the maturity is 6 months. For method 2, the RMSEs are monotonically increasing as T increases. When we extend the maturity from 1 month to 2 months, the RMSE for γ increases very slightly, however the RMSE for α and β increases by 42% and 79% respectively. When further moving to 6 months, the RMSEs shows significant increases, at 90% for α, 137% for β and 34% for γ. From either method, we see that the calibration favors short-term options (with $T = 23$ and 69) than long-term options (with $T = 138$), since the 6-month option sample has the lowest RMSE across all the samples for both methods. In addition, for method 2, we see that the impact of maturities is less significant on γ, but very substantial on α and β. In fact, the RMSEs is tripled for α and quadrupled for β when moving from $T = 23$ to $T = 138$. This suggests that when we filter $h(t)$ from returns, short-term options give lower RMSE than long-term options, and only using long-term options could increase the RMSEs for

α and β by 3-4 times. The probable factor behind this may be that since the prices of long-term options are less dependent on spot variances $h(t)$ and more dependent on its long-run level, when using long-term options, the information captured from the filtered spot variances will not be as large as using short-term options, hence the RMSEs are bigger.

Table A.4: RMSEs for option samples with different variances

	$h(t) = 0.5309e$-4	$h(t) = 1.0617e$-4	$h(t) = 2.1234e$-4	Benchmark
Method 1				
λ	0	0	0	0
ω	0	0	0	0
α	1.2114e-06	1.3700e-06	1.7379e-06	3.8133e-07
β	0.1485	0.1465	0.1987	0.0125
γ	47.4183	387.0449	423.4044	16.6518

Lastly, we present the comparison of RMSEs for different $h(t)$. Note that the benchmark option sample uses the simulated true variances $h(1), ..., h(250)$ as the variances for our option sample. To control the volatility level, we simply fix the conditional variance at the 0.5×, 1× and 2× of long-run level for all options, and only use Options-only calibration method 1 to compare the RMSEs, because the variances cannot be filtered from returns. Table A.4 shows the RMSEs for different level of variances. We note that, although the RMSEs show an increasing pattern (except only for β from $h(t) = 0.5309e$-4 to $h(t) = 1.0617e$-4), the numbers are much larger than the benchmark sample, due to the fact that Options-only calibration method 1 fails to identify individual parameters. From this table, we might conclude that small variances have lower RMSEs for all parameters compared to larger variances. Nonetheless, because the RMSEs when restricting variances to be a constant in each sample are much greater than the benchmark, the numbers tell that we should use a variety option sample with a spread level of variances, instead of keeping the level consistent.

Summary

From the above comparison, we can reach a few positive conclusions. First, the Options-only calibration method 2 is superior to method 1, as the volatility filtering helps eliminate absurd calibration results. When focusing on method 2 specifically, we see that the calibrated model parameters have lower RMSEs with out-of-money options compared to in- and at-the-money options. In terms of time-to-maturity, the RMSEs for all parameters are lower with the short maturity sample, than the sample with longer maturities. Both findings show the importance of using out-of-money and short maturity options in calibrating the model. The evidence also appears when the benchmark option sample has a very low level of RMSEs. In the comparison, only the sample with strike prices fixed at $K = 110$ has an RMSE of 12.370 for γ, which is lower than the benchmark sample at 12.847. For α and β, only the option sample with time-to-maturity fixed at $T = 23$ has lower RMSEs, at 3.593e-7 and 0.011 respectively, compared to the benchmark group at 3.872e-7 and 0.012 respectively. While the superiority of the benchmark group is established because it mostly contains out-of-money and short maturity options, it is also demonstrated that one should use a complete cross-section of options in calibrating the model. While it is better to include out-of-money and short maturity options, the calibration also benefits when the sample contains a large variety of options with different time-to-maturity, strike prices and spot variances.

Appendix B

Calibration result with C-H-J 2006 parameters

In this appendix, we include the Joint estimation-calibration results when the C-H-J 2006 parameters are used when generating the data. The C-H-J 2006 parameters behave similarly, compared to the C-H-J 2013 ones, and the approach to fix $\omega = 0$ is used as well for this exercise. Therefore, as we will see, similar results also hold. The purpose to report the numbers for the new set of parameters is twofold. First, the numbers will demonstrate the validity of our previous conclusions, since different sets of parameters have the same result. Second, the result shows the consequence when the true ω is non-zero, but we fit the model with the (misspecified) restriction of $\omega = 0$. It turns out that such misspecification does not hurt the experiment nor the result.

Similar to Chapter 4, we present the table for the Joint calibration-estimation for various N, and also the block of RMSEs, as below. The MLE results in this appendix are obtained with `fminunc`. We ran each estimation with 2 very different initial values, and the optimization results are exactly the same, hence validated our results that we obtained the global optimum.

Table B.1: Estimation result: Joint estimation

Parameter	True value	$N_r = N_o = 500$		$N_r = N_o = 1500$		$N_r = N_o = 4500$	
		MLE (SE)	Sample SE	MLE (SE)	Sample SE	MLE (SE)	Sample SE
λ	2.772	12.823 (5.146)	4.517	-1.758 (2.394)	2.422	2.804 (1.495)	1.390
ω	3.038e-9	0 -	-	0 -	-	0 -	-
α	3.66e-6	3.974e-6 (7.190e-7)	7.005e-7	2.954e-6 (3.245e-7)	4.699e-7	3.529e-6 (2.267e-7)	2.355e-7
β	0.9026	0.901 (1.504e-2)	1.428e-2	0.914 (0.721e-2)	0.887e-2	0.909 (0.490e-2)	4.467e-3
γ	128.4	111.42 (13.282)	19.916	143.89 (10.104)	9.647	124.84 (5.073)	5.596
$\beta + \alpha\gamma^2$	0.965	0.950	0.778e-2	0.975	0.520e-2	0.964	0.277e-2
Annualized Vol.	15.41%	14.19%	0.988%	17.21%	0.547%	15.68%	0.298%

The above table shows the joint calibration results with different number of returns and options using the Gaussian joint likelihood (3.23) as the objective function. The N_r and N_o are the number of returns and options in the sample, respectively. We increase the number of returns and options without changing their ratio in the sample. The option data sets are generated based on the true parameters (C-H-J 2006 parameters, see (2.16) and (2.17)) and with the specification in (4.10). The calibration is conducted with ω restricted to 0, to enhance the robustness of the numerical optimization (this is also applied by Christoffersen, Heston and Jacobs (2013)). Note that the risk-neutral parameter $\gamma^* = \gamma + \lambda + 0.5$ is used in finding option prices, where 2.772 is the true λ. The estimated standard errors are computed via the outer product of the gradient and is shown in parenthesis. The sample standard error is computed by sample standard deviation of the optimized parameters for 100 simulated samples for each N. The persistence $(\beta + \alpha\gamma^2)$ and annualized volatility (computed via (2.10)) are computed at MLE values.

Table B.2: Comparison of the maximum likelihood estimator's sample RMSE (of 100 samples) for different number of returns and options in joint estimation. Data is simulated based on C-H-J 2006 parameters. Identical random seeds are used (the seed for small samples are nested in big samples). The standard deviation of vega noise is 4.96%.

Sample RMSE	No. of options	0	500	1500	4500
No. of returns					
500	λ	4.6540	4.4951	4.5766	4.4989
	α	1.5606e-06	6.9711e-07	4.8890e-07	2.9249e-07
	β	6.6727e-02	1.4326e-02	9.6849e-03	5.6824e-03
	γ	131.8441	20.1938	10.0499	7.5250
1500	λ	2.5055	2.4137	2.4098	2.4093
	α	7.8786e-07	5.9028e-07	4.7036e-07	2.9207e-07
	β	1.5387e-02	1.3245e-02	9.0844e-03	5.4584e-03
	γ	28.1684	15.0690	9.7247	6.4416
4500	λ	1.3885	1.3770	1.3777	1.3887
	α	4.6034e-07	3.8791e-07	3.3864e-07	2.3433e-07
	β	9.1951e-03	8.3457e-03	6.7347e-03	4.4522e-03
	γ	14.9860	11.4223	8.1858	5.5898

Appendix C

Approaches to incorporate the calibration with ω

The numerical calibration with ω has been a problem when we bring option prices into the calibration. This is because the objective function f is not defined when $\omega < 0$. In particular, when we use an unconstrained optimization algorithm, because the true ω is too close to the unacceptable region, the optimization algorithm will step into the unacceptable region and hence break down the process. The widely applied approach to this problem is to fix ω to be 0, as seen in Christoffersen, Heston and Jacobs (2013). However, because the in-sample model fitting with the restricted model is always worse than the full model, we always want to bring ω into the calibration as well. During our empirical studies, we find the uses of constrained optimization algorithms are not reliable. Overall, they can produce an optimized result, but the result gives a lower log-likelihood than when we used to do unconstrained optimizations with ω fixed to 0. Thus, in this appendix, we briefly outline some numerical tricks that can be applied to this problem when we stick with unconstrained optimizations, and state their effectiveness. The following numerical tricks try to extend the objective function to make it defined on the whole Euclidean space.

- Trick 1: Fix $f(\omega) = f(0)$ when $\omega < 0$. This method works theoretically, however, in reality, the optimization will still go to the $\omega < 0$ region regardless of the initial value, and be stuck in there. Hence, the objective of effectively estimate ω cannot be achieved in 90% of the cases.

- Trick 2: Define $f(\omega) = f(-\omega)$ when $\omega < 0$, where f is the objective function. This is a fix to the previous method and indeed we shall see a meaningful ω estimation. In a preliminary test of 20 samples, 13 of which showed higher log-likelihood when this trick is applied, compared to the benchmark ($\omega = 0$ fixed) approach.

- Trick 3: Use an exponential link function, so that the objective function becomes $f(\exp(\omega))$. However, this method is highly influenced by how we scale the parameters. Because of the exponential link function, if we give ω a larger scale, then ω will dominate other parameters in the numerical optimization process. If we give ω a small scale, then ω does not move far from the initial value. From our previous experiments, the optimization achieved higher log-likelihood values when restricting ω to 0. This suggests that this approach, when applied, needs to be combined with other transformations to achieve meaningful results.[1]

[1] An example of such transformations include estimating (λ, long-run variance (using $\exp(\omega)$), persistence, α, and γ), instead of the usual $(\lambda, \omega, \alpha, \beta, \gamma)$.