OPTION PRICING ACCURACY FOR ESTIMATED STOCHASTIC VOLATILITY MODELS

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yutheeka Gadhyan

August 2010

Yutheeka Gadhyan

APPROVED:

Dr. Robert Azencott
Chairman

Dr. Roland Glowinski
Co-Chairman

Dr. Gemunu Gunaratne

Dr. Tsorng-Whay Pan

Dr. Ilya Timofeyev

ACKNOWLEDGMENTS

I extend my deepest gratitude to my adviser Professor Robert Azencott without whose guidance this thesis would not have been possible. I thank him for his support, encouragement, and patience. His breadth of knowledge, deep insights, energy, and enthusiasm for Mathematics have been both mesmerizing and inspiring. I consider myself fortunate for having worked under the scientific direction of a wonderful teacher and mentor.

I am also deeply indebted to my co-adviser Professor Roland Glowinski for giving me the freedom to pursue my own mathematical interests while providing constant support and invaluable advice.

I owe many thanks to Professors Gemunu Gunaratne, Tsorng W. Pan, and Ilya Timofeyev for taking the time to be a part of my thesis committee and for several valuable suggestions on my dissertation.

I truly appreciate the financial support provided by the Mathematics department for my graduate studies. I thank the department faculty and staff members for always being very helpful and approachable. I am grateful to my teachers at the Mathematical Sciences Foundation, Delhi, for guiding me towards research in Mathematics. I would like to thank my colleagues in the Mathematics department for making my PhD experience happy and stimulating, especially Arjun for several interesting discussions on my research and Mathematics in general.

My parents Mira and Arun Gadhyan have been my strength and inspiration. I can never thank them and my brother Rahul Gadhyan enough for the joy and happiness they bring to my life.

OPTION PRICING ACCURACY FOR ESTIMATED STOCHASTIC VOLATILITY MODELS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yutheeka Gadhyan

August 2010

ABSTRACT

Stochastic differential equations (SDEs) with space dependent coefficients are used to model asset price and volatility dynamics in the mathematical study of financial markets. The valuation of option contracts based on a given asset is then derived by numerically solving associated partial differential equations (PDEs), derived themselves from the SDEs driving the underlying asset and volatility. The parameters of these SDEs have to be estimated from observed data recording daily asset price and volatility. We quantify how the unavoidable errors in the estimation of model parameters impact the valuation of option contracts, generating errors in option pricing.

We have developed and numerically implemented a fast and efficient approximate maximum likelihood approach to estimate the 5 parameters of the widely used Heston model which involves two interacting SDEs from realistic numbers of recorded asset price and volatility data. We study and prove the asymptotic consistency of these estimators and evaluate explicitly their speed of convergence as the number of observations increases. We evaluate the sensitivity of option pricing to parameter estimation errors by solving numerically the 6 PDEs satisfied by the option price and by its derivatives with respect to the 5 underlying parameters. We study the size of the option pricing errors by intensive simulations of SDEs for realistic bench test groups of 5 parameters. We successfully apply our methods to the computation of option pricing errors for actual stock market data : the daily S&P 500 index and its approximate volatility (namely the VIX index).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Stochastic models in finance

Stochastic differential equations (SDEs) with space dependent coefficients have been used to model asset price and volatility dynamics in the mathematical study of financial markets. A widely applied model that captures the joint dynamics of the asset price and its volatility is the Heston stochastic volatility model introduced by Heston in 1993. The Heston model is given by a pair of correlated autonomous stochastic differential equations and depends on five intrinsic parameters. We will consider as our underlying model the joint SDE parametric model of Heston. The popularity of the Heston model in practice is due to it's ability to capture empirically observed stylized asset price and asset volatility behavior and for the numerical tractability of option price formulas under this model.

1.2 Estimation: Heston model

We develop a pragmatic approach to estimate the parameters of Heston's stochastic volatility model with the objective of option pricing under this model. The joint density of the equations is not known and hence the exact likelihood function is not known. We develop

a constrained approximate maximum likelihood method based on an Euler-Maruyama discretization of the Heston SDEs. We estimate the five model parameters under pragmatic constraints. We propose a decoupled estimation of the model and obtain closed form expressions for all the five model parameters. The estimation method is computationally inexpensive. We report small sample results for realistic values of the observation size given the S&P 500 daily data with its approximate volatility, namely the VIX index. We prove the consistency of the estimators and illustrate our theoretical results with detailed numerical examples. For the pricing of options under the principles of *no arbitrage* in the market, an extra parameter called the *market price of volatility risk* needs to be estimated from market data. We propose a first attempt at estimating this parameter from options data.

1.3 Sensitivity: Heston model

Financial contracts such as options were introduced to be used as hedging tools, and are now heavily traded in the market. Trading prices of options are indicators of market expectations for the near future. Due to high volatilities, efficient hedging tools require accurate pricing of the options contract. When the underlying *asset* price, on which the option contract is defined, satisfies a stochastic differential equation, the option price satisfies an associated parabolic partial differential equation whose coefficients are the parameters of the underlying asset price stochastic model. For our purpose, we will think of an asset to be a financial stock and we will use the two terms interchangeably. The true values of the model parameters are never known. To evaluate option prices through robust model based inference from asset dynamics data, it is therefore crucial to understand the impact of parameter estimation errors on the option price. Since in practice the model parameters are estimated from only a small data sample the estimated parameters are indeed fraught with errors. Our goal is to understand how errors in the estimated parameters impact the price of options. Toward this end we develop a novel approach to study the sensitivity of an European option price

with respect to each of its parameters when the underlying asset price and its volatility satisfy Heston's model. We solve numerically the partial differential equation satisfied by the European option price and the partial differential equations satisfied by its derivatives with respect to the parameters of the Heston model. We define and illustrate the impact of option pricing errors due to model estimation with detailed numerical examples for actively traded options on the S&P 500 index.

One possible application of this study is to determine the parameters to which the option price is highly sensitive. This can be used as a guidance in determining the accuracy with which these parameters should be estimated. Clearly high sensitivity parameters will need to be estimated with sharp accuracy while parameters for which the sensitivity is not significant may not require high accuracy estimators.

Another area of application is portfolio hedging. It is often desirable in practice to construct portfolios which are neutral or insensitive to changes in one or more underlying parameter values. For example, in option pricing under the Black-Scholes model assumption the key parameter is the volatility, σ, of the asset price. Under this model, let $P(t, s)$ be the price of a portfolio at time t, consisting of one option where s denotes the value of the underlying asset. This portfolio can be 'vega-hedged' or be made volatility-neutral by adding x units of the underlying asset so that the derivative of $P(t, s) + x.s$ with respect to volatility is zero. The option price $P(t, s)$ under Black-Scholes model is differentiable with respect to volatility. Taking the derivative of the portfolio with respect to volatility and setting it to zero gives that at time t, x should be equal to $-\frac{\partial_\sigma P}{\partial_\sigma s}$. Note that the numerator is just the derivative of the option price with respect to σ. For the Black-Scholes model, the numerator and denominator are known in closed form. For the more realistic stochastic volatility models that we study closed form expressions for the option price derivatives are not known. Our study can therefore be used in the direction of portfolio hedging when

asset prices follow stochastic volatility models.

1.4 Outline of thesis

The detailed structure of the thesis is as follows. Chapter 2 is a brief review of stochastic models used in finance, in particular stochastic volatility models where the volatility of the asset price is a stochastic process. We give references to some papers from the existing literature on the estimation of Heston's model and on the estimation of asset price volatility. In Chapter 3, we describe our constrained approximate maximum likelihood method and give closed form expressions for all the five Heston model parameters. In the Heston model, there is a non-linear interplay between the choice of time unit and the model parameters. We clarify this explicitly by re-parametrizing the Heston model in Section 3.3. In Section 3.3 we define the decoupled estimation strategy and perform a constrained optimization to obtain in closed form the parameter estimators corresponding to the stochastic differential equation driving the asset price volatility. In Section 3.4.4 we obtain the estimator for the asset price process and finally in Section 3.4.5 we estimate the coefficient of correlation between the two processes. In Chapter 4, we prove the consistency of the estimators. The ergodicity property of the continuous process is not preserved after discretization. We therefore study the limits of the estimators as the time between observations goes to zero and the time till which the data is observed goes to infinity. Chapter 3 and Chapter 4 are extensions of the work presented at the 7th AIMS Conference on Dynamical Systems and Differential Equations, Arlington, Texas, May 18-21 2008. In Chapter 5 we present numerical examples for eight benchmark Heston models to validate the consistency of the estimators and the decay in the variance of estimators for different values of the T, the time between observations and S, the time till which the the data is observed. In Section 5.3 we present small sample bias and errors for the estimators corresponding to the S&P 500 and VIX daily data from Jan 03 1006 to Dec 29 2006. The parameter estimation and error

computations can be similarly carried out for the Heston model fitted to intra-day data after considering the implicit behavior of intra-day data.

In Chapter 6 we study the partial differential equation satisfied by the European option price under Heston's model and discuss the existence and uniqueness of the solution. We present a methodology to estimate the market price of volatility risk from observed option price data. The error in the option price is defined as the L_2 norm of the difference between the option price at the true parameter value and the price at the estimator. In Chapter 7 an expression for the L_2 error in the option price is obtained in terms of the L_2 errors in the estimators and the derivative of the option price with respect to the parameters. The sensitivity of the option price to each parameter is defined in Section 7.6. In Section 8.1 the numerical scheme used to solve the six partial differential equations for the option price and its derivatives is described. In Section 8.2 the impact of estimation errors on option price and the sensitivity of the option price to the five underlying parameters is illustrated by considering four different options each for four Heston models in a moderately large neighborhood of the 2006 S&P 500 model. Finally we conclude in Section 8.3 with some applications of our work. Parts of the work in Chapters 6, 7, and 8 was presented at the SIAM Conference Mathematics for Industry: Challenges and Frontiers, San Francisco, California, October 9-10, 2009 and appeared in [7].

Chapter 2

Stochastic models in finance and applications

2.1 Brief history and references

In the theory of continuous time option pricing, the underlying traded or non-traded instrument, on which an option contract is defined, is modeled by a Stochastic Differential Equation. The earliest work in this direction is by Bachelier in his thesis [8] in 1900 where he derived option pricing formulas under the assumption that stock price fluctuations are normally distributed random variables. The asset price model of Bachelier was developed by Samuelson [66] in 1965 who proposed that asset prices follow a *geometric Brownian motion*. The theory of option pricing when the price of an asset is a geometric Brownian motion was developed by Black and Scholes [15] and Merton [58] in 1973. The option pricing theory of Black and Scholes has been popular in practice and significantly extended since then (see e.g., [12, 24, 9]). A key assumption in the Black-Scholes model is that the instantaneous asset price volatility is constant in time irrespective of the direction of the asset price.

The constant volatility assumption of the Black-Scholes model has been called into question by evidence from market data for asset price (e.g., [31, 55, 50, 3]). Black [14] commented on the non-Gaussian distribution of the returns process by analyzing the time series of stock price returns. Empirical evidence from option price data also shows that volatility is not constant [10, 64, 22, 74]. These papers show that the option price varies as a function of the strike price and time-to-maturity of the options contract. The volatility recovered from option price with varying strike price by inverting the Black-Scholes formula (called Black-Scholes implied volatility) is a U-shaped curve called the smile curve in finance jargon [26]. The Black-Scholes option price is an increasing function of volatility and hence such an inversion is possible. As a response, non constant volatility models have been proposed, e.g., local volatility models where volatility is a function of the underlying asset and time [26, 29, 65], and stochastic volatility models where volatility is itself driven by a stochastic process.

2.2 Stochastic models for asset prices and asset volatilities

Stochastic volatility models provide a natural generalization of the Black-Scholes model that describe a more complex market and capture some of the empirical features of the joint series of the asset price and options data (see e.g., [38, 11, 56, 57]). Stochastic volatility models are often given by a system of two possibly correlated stochastic differential equations for the asset price and volatility respectively, (see [47, 67, 73, 70, 44]). In particular, Stein and Stein [70] specify volatility to be given by an uncorrelated arithmetic Ornstein Uhlenbeck process while Heston [44] considers that the square of volatility satisfies the mean reverting square root process of Feller [32].

2.3 Heston model: description and history

The stochastic volatility model of Heston for asset price X_t and squared-volatility σ_t^2 was introduced in 1993 in [44]. In Heston's framework the asset price and volatility satisfy the following pair of stochastic differential equations, (see 3.2),

$$dX_t = \mu X_t dt + \sigma_t X_t dZ_t, \tag{2.1}$$

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \gamma\sigma_t dB_t, \tag{2.2}$$

where Z and B are correlated standard Brownian motions with correlation coefficient ρ. The five parameters in Heston's model allow for greater flexibility in capturing stylized asset price and volatility behavior which cannot be captured by more limited models. For example, empirical evidence in the econometrics literature shows that there is indeed a correlation between the price of an asset and its volatility [14]. In fact, it has been suggested that the asset price is negatively correlated with its volatility in [14]. We find evidence to support this from our numerical results on the S&P 500 data. The Black-Scholes implied volatility surface under Heston's model is in close agreement with empirically observed implied volatility. The option pricing formulas are numerically tractable for the Heston model. Heston [44] provides a semi-closed form for the price of an option in terms of the inverse characteristic function of the underlying probabilities. Existing literature on the numerical methods for option pricing under Heston's model is discussed briefly in 6.

Here we will consider Heston's model as the underlying asset price and volatility model and study first the estimation problem for this model and then the impact of estimation errors on option price under this model.

2.4 Heston model: survey of previous estimation techniques

Option pricing under Heston model requires an additional parameter λ which arises due to the non-tradability of volatility. The true value of λ is not known and has to be estimated from market data. Therefore one needs to estimate six parameters for an application of the Heston SDEs to model asset price and volatility dynamics. The volatility of an asset is not traded in the market. Therefore a time series of observations for the volatility process is not available unlike for asset price such as a stock price whose trading price at any instant is taken to be its observed value at that time. Recent work on the estimation of Heston model use the joint series of the stock price and options data to estimate the parameters including λ and an estimation or filtration method to recover the underlying volatility.

Chernov and Ghysels [20] apply efficient method of moments to estimate the parameters of the Heston model from the joint series of the option price and stock price data. They simulate the required moments by an Euler-Maruyama discretization of the Heston model. The article by [60] describes a least-squares error fit between the model predicted and observed option price to estimate the model parameters and λ. They test deterministic and stochastic optimization algorithms to solve the resulting non-linear problem. Aït-Sahalia and Kimmel [2] propose a maximum likelihood approach to estimate the Heston model parameters. They use a Hermite approximation of the log-likelihood function. They compare estimation results obtained from using the joint series of stock price and options data to the results obtained when using the joint series of the stock price and a volatility proxy, namely the Black Scholes implied volatility. They show that there is a significant computational gain in using the latter series with no real loss of estimation accuracy except that the asset price and volatility series does not allow the estimation of the parameter λ. A survey of estimation techniques for the Heston model and in general for stochastic volatility models is in [19, 38, 16].

In our numerical examples with real data, namely the S&P 500 index; we will use the VIX index as an approximation of volatility. The VIX index is an adjusted Black-Scholes implied volatility computed from a group of 30 options (see [72]). However, our estimation method does not make any assumptions on how the volatility series is obtained. Methods for estimation of the diffusion coefficient from discrete observations have been proposed in the classical literature [37, 27, 25]. We will only assume that an estimate of the volatility series is available. Using the joint asset price and volatility series we will obtain the parameters of the Heston model. We will then estimate separately the parameter λ from options data (see chapter 6).

Chapter 3

Parameter estimation for Heston model : Theoretical construction

3.1 Introduction

We compute fast and robust constrained approximate maximum likelihood estimators for the parameters of the prototypical stochastic volatility model of Heston given the joint series for the asset price and volatility. The Heston model is described by a two-dimensional stochastic differential equation (SDE) where the asset price and its volatility are driven by correlated Ito processes [48]. In this chapter we derive closed form expressions for the parameter estimators, enabling very efficient numerical implementations. We study the asymptotic properties of the estimators as the observation time goes to infinity and the time between consecutive observations goes to zero in chapter 4. We prove that the estimators are consistent and validate numerically that the variance of the estimators goes to zero as the observation time increases. In chapter 5 we illustrate our method by application to actual financial data (S&P 500 index paired with its approximate volatility, namely the VIX index).

3.2 Heston Stochastic Volatility model

Let $(\Omega, \mathcal{F}_t, Pr)$ be the underlying probability space, endowed with a filtration $\mathcal{F}_t$ [21]. We will denote by $X = \{X_t\}_{0 \leq t < \infty}$ the asset price process. Then X_t is the price of the asset at time t, $0 \leq t < \infty$. Let σ_t be the instantaneous relative volatility of the returns process dX_t. We will refer to σ_t as the volatility variable. Let $Y_t = \sigma_t^2$. Then Y_t is the squared volatility at time t and $Y = \{Y_t\}_{0 \leq t < \infty}$ is the associated process. In the Heston model [44], the pair $\{X_t, Y_t\}$ observed in the market satisfies the following coupled stochastic differential equations (SDEs),

$$dX_t = \mu X_t dt + \sqrt{Y_t} X_t dZ_t, \tag{3.1}$$

$$dY_t = \kappa(\theta - Y_t)dt + \gamma\sqrt{Y_t} dB_t, \tag{3.2}$$

where the processes $Z = \{Z_t\}_{0 \leq t < \infty}$ and $B = \{B_t\}_{0 \leq t < \infty}$ are standard Brownian motions adapted to $\mathcal{F}_t$ taking values in $\mathbb{R}$ with $E[dZ_t dB_t] = \rho dt$. We observe that the drift and diffusion coefficients of the SDEs (3.1)-(3.2) are space dependent and autonomous. The diffusion coefficients of (3.1)-(3.2) do not satisfy a global Lipschitz condition and hence the classical existence and uniqueness results for SDEs [39] cannot be applied. The existence and uniqueness of a strong solution for the SDEs (3.1)-(3.2) follows from a result due to Yamada and Watanabe [75] which requires only continuity of the drift coefficient and replaces the Lipschitz continuity constant by a suitably integrable increasing function.

For $\kappa, \theta > 0$ the squared volatility process $\{Y_t\}$ follows the mean reverting square root process studied by Feller [32] and originally used by Cox, Ingersoll, and Ross to model short-term interest rates [23]. The parameter γ is the instantaneous relative volatility of Y_t and ρ is the instantaneous correlation coefficient between the two Brownian motions driving the SDEs. The parameter μ is the mean rate of return of the asset price. The reversion of the volatility process about θ is illustrated in Fig. 3.1 for two different values of κ. The two

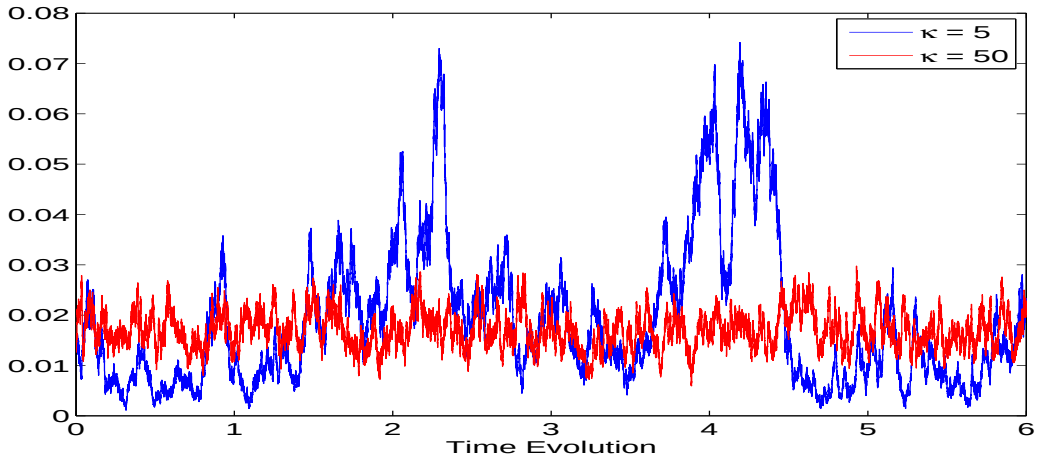


Figure 3.1: Reversion of volatility about its long run mean $\theta = .017$

trajectories are generated by numerical simulation of the SDE (3.2) for $\theta = .017$, $\gamma = .28$ and $\kappa = 5$ and 50.

The singularity of the diffusion coefficient in (3.2) at the origin implies that starting with a non-negative initial value $Y_0 = y_0 \geq 0$, the process can subsequently never take negative values. This follows from the comparison theorem for one-dimensional Ito processes [48]. The classical condition, $2\kappa\theta \geq \gamma^2$ due to Feller [32], ensures inaccessibility of the origin, hence for $Y_0 > 0$, the process $\{Y_t\}$ remains strictly positive for all t, $Pr - almost\ surely$. This condition is an important restriction if we want Y_t to realistically model the square of asset price volatility. The properties of SDE (3.2) by itself are of interest because as we will see later we decouple the estimation of (κ, θ, γ) from the estimation of μ and ρ. The law of Y_s conditional on its value $Y_t = y_t$ at time t $(t < s)$ is a scaled non-central chi-square with $4\kappa\theta/\gamma^2$ degrees of freedom and parameter of non-centrality $2y_t e^{-(s-t)}$ [23].

Proposition 3.2.1. *The first two moments $E[Y_t]$ and $E[Y_t^2]$ of Y_t satisfying the SDE* (3.2) *are given by,*

$$E[Y_t] = E[Y_0]e^{-\kappa t} + \theta(1 - e^{-\kappa t}),$$

and

$$E[Y_t^2] = e^{-2\kappa t}(Y_o^2 + (\frac{2\kappa\theta+\gamma^2}{\kappa})(\frac{\theta}{2} - Y_o)) + e^{-\kappa t}(\frac{2\kappa\theta+\gamma^2}{\kappa})(Y_o - \theta) + \theta\frac{2\kappa\theta+\gamma^2}{2\kappa}.$$

Proof. Writing SDE (3.2) in the integral form we get,

$$\begin{aligned}
&\int_0^t dY_u = \int_0^t \kappa(\theta - Y_u)du + \gamma\int_0^t \sqrt{Y_u}dZ(u),\\
&\Rightarrow E[Y_t] = E[Y_0] + \int_0^t E[\kappa(\theta - Y_u)]du,
\end{aligned}$$

To obtain the preceding equation we observe that the stochastic integral $\int_0^t \sqrt{Y_u}dZ_u$ is a continuous $\mathcal{F}_t$ martingale with $E[\int_0^t \sqrt{Y_u}dZ_u] = 0$ because $\int_0^t E[Y_u]du < \infty$ on any finite interval $(0,t)$, by the definition of stochastic integrals [21]. Differentiating both sides with respect to the time variable t we obtain

$$\frac{d}{dt}E[Y_t] = E[\kappa(\theta - Y_t)] = \kappa\theta - \kappa E[Y_t].$$

Denoting $E[Y_t]$ by $f(t)$ we get the following ordinary differential equation (ODE) from the preceding equation,

$$f'(s) = \kappa(\theta - f(s)).$$

Integrating both sides from 0 to t we get,

$$\begin{aligned}
\kappa t &= \ln(\theta - f(0)) - \ln(\theta - f(t)),\\
\Rightarrow E[Y_t] &= E[Y_0]e^{-\kappa t} + \theta(1 - e^{-\kappa t}).
\end{aligned}$$

Similarly we can compute the second moment of Y_t,

$$E[Y_t^2] = e^{-2\kappa t}(Y_o^2 + (\frac{2\kappa\theta+\gamma^2}{\kappa})(\frac{\theta}{2} - Y_o)) + e^{-\kappa t}(\frac{2\kappa\theta+\gamma^2}{\kappa})(Y_o - \theta) + \theta\frac{2\kappa\theta+\gamma^2}{2\kappa}.$$

□

We observe from Proposition 3.2.1 that $E[Y_t] \to \theta$ and $Var(Y_t) \to \frac{\gamma^2\theta}{2\kappa}$ as $t \to \infty$ at an exponential speed with rate κ. As noted in [23], for $\kappa, \theta > 0$, the process $\{Y_t\}$ has a stationary Gamma distribution [32], with density $g(y)$,

$$g(y) = \frac{\xi}{\Gamma(\nu)}(\xi y)^{\nu-1}e^{-\xi y}\mathbf{1}_{\{y>0\}},$$

where

$$\xi = v/w = 2\kappa/\gamma^2, \quad \nu = u/w = 2\kappa\theta/\gamma^2, \quad \Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt.$$

Denote by H the vector of five parameters determining a Heston model,

$$H = \{\kappa, \theta, \gamma, \rho, \mu\}.$$

The natural domain in $\mathbb{R}^5$ for the unknown vector of parameters is defined by the following classical constraints,

$$\kappa > 0,\, \theta > 0,\, \gamma > 0, \quad 0 < \theta < 1, \quad -1 \leq \rho \leq 1, \quad 2\kappa\theta > \gamma^2. \tag{3.3}$$

The condition $\theta < 1$ ensures a realistic upper bound on its estimated value. The volatility $\sqrt{Y_t}$ in (3.1) is expressed relative to the asset price. For instance in the historical data for the S&P 500, even in extremely volatile markets, such as the period between 2006 and 2008, the implied volatility was not higher than 85% [18]. The parameter θ is the long run mean of Y_t and should therefore realistically stay under 1. The parameter γ is taken to be positive by convention. We impose the strict constraint $2\kappa\theta > \gamma^2$ because a closed form expression for the first moment of the process $\{1/Y_t\}$ with respect to the stationary distribution exists under this condition (see chapter 4). We note that under the condition $4\kappa\theta = \gamma^2$, the

process $\{\sigma_t\} = \{\sqrt{Y_t}\}$ is the well known Ornstein-Uhlenbeck stochastic process [35]. The proof follows from a direct application of Itô's lemma (6.6).

3.3 Estimation of parameters for Heston joint SDEs

3.3.1 Estimation strategy

In practice, volatility is not directly observed, but as discussed in the last chapter, well known methods are available to estimate these volatility values from market data. Hence we consider that we are given a data set of N joint observations for the asset price X_t and its volatility $\sqrt{Y_t}$, and we want to estimate the vector of model parameters $H \in \mathbb{R}^5$.

We will use a maximum likelihood estimation approach to estimate the parameters after discretization of the joint SDEs model. The joint density of $\{X_t, Y_t\}$ is not available in closed form, so that an explicit compact expression for the log-likelihood function of our joint SDEs (3.1)-(3.2) is not available. We will therefore compute the log-likelihood function from the discretized model. The parameters κ, θ and γ appear only in SDE (3.2) while the parameter μ appears only in SDE (3.1). After a time discretization of the SDEs (3.1)-(3.2), we will naturally decouple the estimation of (κ, θ, γ) and (μ, ρ). The maximum likelihood estimators (MLE) of (κ, θ, γ) will be based on $\{Y_t\}$ only. The parameter μ is then estimated by MLE based on the first SDE only. Finally we will recover an explicit approximation of the two underlying correlated Brownian motions $\{Z_t\}$ and $\{B_t\}$ and then estimate ρ as their empirical correlation coefficient.

Decoupling the estimation of parameters provides a strong numerical gain, because our parameter estimators can be explicitly computed and we can prove consistency of our estimators.

3.3.2 Change of time unit in the joint Heston SDEs

In the Heston model, there is a non-linear interplay between the choice of time unit and the model parameters, and one needs to clarify explicitly this relation, since in practical model fitting to discrete market data, the time unit is somewhat arbitrary.

Proposition 3.3.1. *Let $\{X_t, Y_t\}$ be solutions of the joint SDEs* (3.1)-(3.2) *above. Fix a new "time unit" $T > 0$, and define $U_s = X_{sT}$ and $V_s = Y_{sT}$ for all $s > 0$. Then the process trajectories $\{U_s, V_s\}$ have the same joint probability distributions as the solutions of the joint SDEs*

$$dU_s = T\mu U_s ds + \sqrt{T}\sqrt{V_s}U_s dZ_1(s), \tag{3.4}$$

$$dV_s = T\kappa(\theta - V_s)ds + \gamma\sqrt{TV_s}dB_1(s), \tag{3.5}$$

where $Z_1(s), B_1(s)$ are standard Brownian motions with correlation ρ.

Proof. Define new processes $Z_1(s), B_1(s)$ by,

$$Z_1(s) = \frac{Z_{sT}}{\sqrt{T}}, \quad B_1(s) = \frac{B_{sT}}{\sqrt{T}}.$$

Then the processes $Z_1(s), B_1(s)$ are both Gaussian with mean zero and have the same covariance function $K(s,t) = min(s,t)$. Hence they are Brownian motions, with instantaneous correlation equal to ρ. Starting from the original SDEs (3.1)-(3.2), the effect of our linear time change can now be seen to generate equations (3.4)-(3.5) [4]. □

Note that in practice only discrete observations (daily or intraday) are available. We define here T to be the time separating two consecutive available discrete observations, and we hence assume that T has already been selected and fixed before implementing an estimation procedure. Typically for the modeling of daily observations, one fixes $T = 1/252$.

We shall focus first on the estimation of the 3 parameters (κ, θ, γ), using only a discrete set of N volatility observations $V_n = Y_{nT},\ n = 0, 1, 2, \ldots, N-1$, where Y_t is solution of the SDE (3.2).

3.3.3 Reparametrization of the volatility SDE

As seen above, the new continuous time square volatility process $V_s = Y_{sT}$ is a solution of SDE (3.5), analogous to SDE (3.2), but with new parameters u, v, w given by

$$u = T\kappa\theta, \quad v = T\kappa, \quad w = \frac{T\gamma^2}{2}. \tag{3.6}$$

With these notations, the SDE (3.5) verified by $\{V_t\}$ can clearly be written as

$$dV_s = (u - vV_s)ds + \sqrt{2w}\sqrt{V_s}dB_1(s). \tag{3.7}$$

Note that these 3 new parameters are proportional to the time interval T between two successive observations of $V_n = Y_{nT},\ n = 0, 1, 2, \ldots, N-1$. This change of parameters is clearly invertible by the formulas

$$\kappa = v/T, \quad \theta = u/v, \quad \gamma = \sqrt{2w/T}. \tag{3.8}$$

Note also that the domain of admissibility $\mathcal{C} \subset \mathbb{R}^3$ for the vector of new parameters $P = (u, v, w)$ is a *convex cone* immediately deduced from the set of constraints (3.3), namely

$$\mathcal{C} = \{(u, v, w) : 0 < w < u < v\}. \tag{3.9}$$

We will now define and study *nearly optimal estimators* $\hat{P} = (\hat{u}, \hat{v}, \hat{w})$ of the unknown vector $P \in \mathcal{C}$ given a set of discrete observations V_n of the process $V_s = Y_{sT}$ driven by the SDE (3.7). Further below, we will then generate estimators $(\hat{\kappa}, \hat{\theta}, \hat{\gamma})$ of (κ, θ, γ) by the

deterministic transformation (3.8)

$$\hat{\kappa} = \hat{v}/T, \quad \hat{\theta} = \hat{u}/\hat{v}, \quad \hat{\gamma} = \sqrt{2\hat{w}/T}. \tag{3.10}$$

3.4 Maximum likelihood estimators for volatility parameters

3.4.1 Euler discretization of the volatility process

For $n = 0, 1, \ldots, N$, let $V_n = Y_{nT}$ be the $N + 1$ observed data for the squared volatility process $\{Y_t\}$ where T is the known and fixed time between consecutive observations. We obtain closed form expressions for the parameter estimators such that the Euler discretization of (3.7) best fits the given data in a maximum likelihood sense. The classical Euler discretization of SDE (3.7) replaces the SDE by the approximate recurrence relations

$$V_{n+1} \approx V_n + u - vV_n + \sqrt{2w}\sqrt{V_n}\Delta B_1(n), \quad n = 0, 1, 2, \ldots, N-1, \quad V_0 = Y_0 = y_0 > 0, \tag{3.11}$$

where $\Delta B_1(n) = B_1(n+1) - B_1(n)$ and $y_0 > 0$ is a given fixed constant.

The convergence of the Euler discretization scheme for the volatility dynamics as $T \to 0$ has been studied in [45]. The standard convergence theory for numerical simulations of SDEs (see [52]) indeed does not cover the Heston SDEs, since the diffusion coefficient of V_s does not satisfy a global Lipschitz condition. For the Euler discretization of the Heston SDE (3.7), the results of [45] show that, over fixed bounded time intervals $[0, S]$, the Euler discretization converges pathwise to the true process as $T \to 0$. They also show that for any fixed $T < \frac{2}{\kappa}$, the first moments of the discretized process converge to the true moment as $N \to \infty$. Of course for large T the Euler approximation becomes a very fuzzy approximation of the true continuous dynamics. In our numerical study, we use daily observations spanning one year for the asset price and volatility, and in agreement with current practice, we select

the fairly small value $T = 1/252$.

3.4.2 Approximate maximum likelihood for the volatility process

We now generate estimators of $P = (u, v, w)$ by maximizing the likelihood of the observations under the discretized approximate dynamics (3.11). From (3.11), we derive

$$\sqrt{2w}\Delta B_1(n) \approx Q_n = \frac{(\Delta V_n - u + vV_n)}{\sqrt{V_n}},$$

for $n = 0, 1, \ldots, N-1$ where $\Delta V_n = V_{n+1} - V_n$. Under the discrete approximation, the random variables Q_n are independent and Gaussian with mean zero and variance $2w$. Define

$$S_N = \frac{1}{N}\sum_{n=0}^{N-1} Q_n^2.$$

The sum of squares S_N is clearly a positive quadratic function of u, v, w given by

$$S_N = a + bu + cv + \frac{1}{2}du^2 - 2uv + \frac{1}{2}fv^2,$$

where the statistics a, b, c, d, f are explicit functions of the N observations $\{V_n = Y_{nT},\ 0 \le n \le N-1\}$, namely

$$\begin{aligned} a =& \frac{1}{N}\sum_{n=0}^{N-1}\frac{(\Delta V_n)^2}{V_n},\quad b = \frac{-2}{N}\sum_{n=0}^{N-1}\frac{\Delta V_n}{V_n},\quad c = \frac{2}{N}\sum_{n=0}^{N-1}\Delta V_n, \\ d =& \frac{2}{N}\sum_{n=0}^{N-1}\frac{1}{V_n},\quad f = \frac{2}{N}\sum_{n=0}^{N-1}V_n. \end{aligned} \tag{3.12}$$

We note that we almost surely have

$$a > 0,\ d > 0,\ f > 0,\quad df - 4 > 0,\quad 2af - c^2 > 0,\quad d + f - 4 > 0. \tag{3.13}$$

Indeed the positivity of a, d, f follows from the almost sure positivity of all $V_n = Y_{nT}$, and the Cauchy-Schwartz inequality in $\mathbb{R}^N$ implies

$$df = \frac{4}{N^2}\sum_{n=0}^{N-1}\frac{1}{V_n}\sum_{n=0}^{N-1}V_n \;>\; 4\frac{1}{N}|\sum_{n=0}^{N-1}\frac{1}{\sqrt{V_n}}\sqrt{V_n}| = 4.$$

Combining this with a generic inequality we have $(d+f)^2 \geq 4df > 16$ and hence $d+f > 4$. Another Cauchy-Schwartz application shows that $2af - c^2 > 0$.

We denote by f the true density function of the continuous process Y_t. By definition Y_t is a Markov process and the Chapman-Kolmogorov equation [63] applied to $V_n = Y_{nT}$ then yields,

$$f(V_0, V_1, V_2, \ldots, V_N) = \Pi_{n=1}^{N} f(V_n|V_{n-1}),$$

for fixed $V_0 = Y_0 = y_0$. With a slight abuse of notation we use V_n to denote a random variable here. From the discrete approximation (3.11) we get,

$$f(V_n|V_{n-1}, P) \approx \frac{1}{\sqrt{4\pi w V_{n-1}}}\exp\{-\frac{1}{4w}(V_n - u - vV_{n-1})^2\}, \quad n = 1, 2, \ldots, N.$$

The *approximate log-likelihood function* $\tilde{L}_N(P)$ of the observations $V_0, V_1, \ldots, V_N$, computed according to the (approximate) discretized dynamics is then given up to a constant by

$$\frac{2}{N}\tilde{L}_N(u, v, w) = -\log 2\pi - \log(2w) - \frac{S_N}{2w}.$$

Our estimator $\hat{P}$ will hence be computed by minimizing, for $P \in \mathcal{C}$ the function

$$L(P) = L_N(P) = -\frac{2}{N}\tilde{L}_N - \log 2\pi = \log(2w) + \frac{S_N}{2w}. \tag{3.14}$$

The estimation problem formulated in terms of the variables u, v, w is then:

$$\begin{aligned} \min\ L(u,v,w) =& \log(2w) + \frac{1}{2w}[a + bu + cv + \frac{1}{2}du^2 - 2uv + \frac{1}{2}fv^2], \\ & \text{on the convex cone } \mathcal{C}: \quad 0 < w < u < v. \end{aligned} \tag{3.15}$$

3.4.3 Constrained optimization of approximate log likelihood

The objective function $L(u, v, w)$ is a differentiable function of its parameters on $\mathcal{C}$ but it is not convex on $\mathcal{C}$, as seen by computing the diagonal of the Hessian H_L of L.

$$H_L = \frac{1}{2w}\begin{pmatrix} d & -2 & -\frac{b-2v+du}{w} \\ -2 & f & -\frac{c-2u+fv}{w} \\ -\frac{b-2v+du}{w} & -\frac{c-2u+fv}{w} & \frac{2}{w^2}(-w+S_N) \end{pmatrix}.$$

The third term on the diagonal of this matrix may not be positive for a fixed value of T and small N. For fixed $w > 0$ however, L is a convex function of (u, v) due to inequalities (3.13). The true parameter vector P is in the open cone $\mathcal{C}$ but the infimum of L on $\mathcal{C}$ may a priori be reached on the closed convex cone

$$\overline{\mathcal{C}} = \mathcal{C} \cup \partial\mathcal{C} = \{(u, v, w) : 0 \leq w \leq u \leq v\}.$$

We clearly have $L(P) \to \infty$ when $P \to \infty$ in $\overline{\mathcal{C}}$, and hence L must actually reach its minimum on $\overline{\mathcal{C}}$.

Let $P = (u, v, w) \in \overline{\mathcal{C}}$ be any minimizer of L on $\overline{\mathcal{C}}$. We distinguish the following four different cases according to the position of the minimizer in $\overline{\mathcal{C}}$.

Case 1: *The minimizer $P = (u, v, w)$ belongs to $\mathcal{C}$:*

Since $\mathcal{C}$ is open the gradient of L must be zero at P. The equations $\partial_u L(p) = \partial_v L(p) = 0$

immediately imply

$$u = -\frac{bf + 2c}{df - 4}, \quad v = -\frac{2b + cd}{df - 4}.$$

Since S_N is positive definite, u, v obtained as above minimize S_N and hence L for fixed w. Since P is in $\mathcal{C}$ we must have $u > v > 0$, which in view of (3.13) is equivalent to the requirement

$$2b + cd < bf + 2c < 0.$$

The quadratic form S_N then takes the value

$$A = S_N(u, v) = a - \frac{1}{2(df - 4)}(b^2 f + 8bc + c^2 d).$$

Note that by its definition the statistic A is a sum of squares and is therefore always positive. To determine w in the open interval $J =]0, u[$, we now minimize

$$B(w) = \log 2w + \frac{A}{2w}.$$

Since w is an interior point of J we have $B'(w) = 0$ and hence $w = A/2$. Since B' is negative on $(0, A/2)$ and positive on $(A/2, \infty)$, we establish that $w = A/2$ is a minimizer of B. To enforce $w \in J$, we must have $A/2 < u$ which is equivalent to

$$2a(df - 4) - (b^2 f + 4bc + c^2 d) + 4(2c + bf) < 0.$$

Hence Case 1 is realized if and only if the following conditions hold

$$2b + cd < bf + 2c < 0 \quad \text{and} \quad 2a(df - 4) - (b^2 f + 4bc + c^2 d) + 4(2c + bf) < 0. \quad (3.16)$$

Then the minimizer $P \in \mathcal{C}$ is unique and given by

$$u = -\frac{bf+2c}{df-4}, \quad v = -\frac{2b+cd}{df-4}, \quad w = \frac{a}{2} - \frac{1}{4(df-4)}(b^2 f + 4bc + c^2 d). \tag{3.17}$$

Note also that as soon as the Case 1 applicability conditions (3.16) are satisfied, the unique minimizer P thus obtained on $\mathcal{C}$ is also necessarily an absolute minimizer on the closure $\overline{\mathcal{C}}$, and hence we do not need to study the 3 other "boundary cases" analyzed below.

Case 2: *Case 1 is not applicable and the minimizer $P = (u, v, w)$ verifies $0 < w = u < v$:* Then for a fixed w, L is a function of v only in an open set. We must then have $\partial_v L(w, v, w) = 0$ which immediately implies

$$v = \frac{-c+2w}{f}.$$

By definition of Case 2 we must have $w < v$, whence the condition $c + (f-2)w < 0$, which is equivalent to imposing a single one of the following three mutually exclusive constraints on w:

$$\begin{aligned} &f < 2 \quad \text{and } w > \max(0, \frac{c}{2-f}), \\ &f > 2 \quad \text{and } c < 0 \text{ and } 0 < w < \frac{c}{2-f}, \\ &f = 2 \quad \text{and } c < 0 \text{ and } w > 0. \end{aligned} \tag{3.18}$$

Then $L(u, v, w) = L(w, \frac{-c+2w}{f}, w)$ becomes a function $G(w)$ of w only given by

$$G(w) = L(w, \frac{-c+2w}{f}, w) = \log 2w + \frac{(2af - c^2) + 2w(bf+2c) + w^2(df-4)}{4fw}.$$

Clearly w must minimize $G(w)$ in w under one of the 3 mutually exclusive constraints (3.18). This forces $G'(w) = 0$ which yields the following quadratic equation for w,

$$4fw^2G' = (df - 4)w^2 + 4fw - (2af - c^2) = 0.$$

Let

$$q(w) = (df - 4)w^2 + 4fw - (2af - c^2) = 0,$$

which in view of (3.13), has a unique positive solution given by

$$w = \frac{-2f + \sqrt{4f^2 + (2af - c^2)(df - 4)}}{df - 4}.$$

It follows that $q(w)$ (and hence also G') is negative on the interval $(0, w)$ and positive on (w, ∞). To compare w to $c/(2 - f)$, we define the statistic

$$R = (2 - f)^2Q(c/(2 - f)) = (df - 4)c^2 + 4fc(2 - f) - (2af - c^2)(2 - f)^2. \qquad (3.19)$$

We now determine the constrained minimizer under the conditions (3.18). We consider separately the three versions of constraint (3.18):

(i) $f < 2$ and $c > 0$: If $R > 0$ so that G reaches its constrained minimum for $w = c/(2 - f)$, which violates the strict constraint (3.18); so this case is not acceptable. If $R < 0$ we must have $0 < c/(2 - f) < w$ so that G reaches its constrained minimum for w.

(ii) $f < 2$ and $c \leq 0$: The interval defined by (3.18) is simply $0 < w$ and we have just seen that G' is negative on $(0, w)$ and positive on (w, ∞), so that the minimum of G is reached at w.

(iii) $f > 2$ and $c \geq 0$: Since $2 - f < 0$, we have $c/(2-f) \leq 0$. Therefore the constraint $0 < w < c/(2-f)$ cannot be satisfied and we don't have an acceptable solution under this condition.

(iv) $f > 2$ and $c < 0$: Then $c/(2-f) > 0$. We study again the sign of R. If $R > 0$ we must have $w < c/(2-f)$ so that G reaches its constrained minimum for w, which satisfies the strict constraint in (3.18). If $R < 0$ we must have $0 < c/(2-f) < w$ so that G reaches its constrained minimum for $w = c/(2-f)$ which does not satisfy the strict constraint in (3.18) and hence this condition is not acceptable.

(v) $f = 2$ and $c < 0$: In this case we solve only under the positivity constraint and the constrained minimizer is w.

This analysis of the feasibility conditions (3.18) shows that Case 2 can only occur if the statistics a, b, c, d, f belong to one of the 4 following mutually exclusive situations:

(1) $f < 2$ and $c > 0$ and $R < 0$.

(2) $f < 2$ and $c \leq 0$.

(3) $f > 2$ and $c < 0$ and $R > 0$.

(4) $f = 2$ and $c < 0$.

Whenever one of these 4 conditions is satisfied, there is a unique minimizer $P = (u, v, w)$ such that $0 < w = u < v$, given by

$$u = w = \frac{-2f + \sqrt{4f^2 + (2af - c^2)(df - 4)}}{df - 4} \quad \text{and} \quad v = \frac{-c + 2w}{f}. \tag{3.20}$$

Case 3: *Case 1 is not applicable and the minimizer $P = (u, v, w)$ verifies $0 < w < u = v$:* On the set $0 < w < u = v$, the function $L(u, v, w)$ becomes a function $H(v, w)$ given by,

$$H(v, w) = L(v, v, w) = \log(2w) + \frac{1}{2w}[a + (b + c)v + (\frac{d}{2} + \frac{f}{2} - 2)v^2].$$

We must then have $\partial_v H = \partial_w H = 0$. These two equalities easily show that the minimizer u, v, w is uniquely determined and given by

$$\begin{aligned} u =& v = -\frac{(b+c)}{(d+f-4)}, \\ w =& \frac{a}{2} - \frac{(b+c)^2}{4(d+f-4)}. \end{aligned} \tag{3.21}$$

Since $d+f-4 > 0$, the minimizer just computed verifies the Case 3 conditions $0 < w < u = v$ if and only if we have the conditions,

$$4(b+c) < (b+c)^2 - 2a(d+f-4) < 0. \tag{3.22}$$

Case 4: *Case 1 is not applicable and the minimizer $P = (u, v, w)$ verifies $u = v = w > 0$:* In this case $L(u, v, w)$ becomes a function $K(w)$ given by

$$K(w) = \log(2w) + \frac{1}{2w}(a + (b+c)w + \frac{1}{2}(d+f-4)w^2).$$

Solving $K'(w) = 0$ we obtain the unique minimizer

$$u = v = w = \frac{a}{1 + \sqrt{1 + a(d+f-4)/2}}, \tag{3.23}$$

which has positive coordinates since $a > 0$ and $d + f - 4 > 0$.

Clearly the last boundary case $u = v = w = 0$ is never a minimizer. Finally we see that the minimization problem (3.15) on the closed $\overline{\mathcal{C}}$ will almost surely have a unique solution $P^* = (u^*, v^*, w^*)$ computed as follows.

Given the discrete volatility observations V_n, we compute the statistics a, b, c, d, f. For each $j = 1, 2, 3, 4$, we shall say that Case j is "applicable" to our data if the explicit

validity conditions given above for Case j are satisfied by a, b, c, d, f, and we then define $L_j = L(P(j))$ where $P(j)$ is the unique minimizer for Case j, explicitly computed above. If Case j is not applicable to our data, we set $L_j = +\infty$.

If Case 1 is applicable, we do not need to consider any other case, as explained above, and we define our estimator $\hat{P}$ of P by $\hat{P} = P(1)$. If Case 1 is not applicable, the index $k = 2, 3, 4$ of the optimal applicable Case k is determined by $L_k = \min(L_2, L_3, L_4)$ and we then define P^* by $P^* = P(k)$. However the parameter vector P^* is then on the boundary $\partial\mathcal{C}$, and boundary vectors are not acceptable parameter vectors.

We shall see in chapter 4 that for small T and large fixed $S = NT$ this situation has a very small probability of occurrence, which indeed tends to zero as $T \to 0$ and $S = NT \to \infty$. Our approximate maximum likelihood estimator $\hat{P}$ for the unknown vector of parameters P will then be any $\hat{P} = (\hat{u}, \hat{v}, \hat{w}) \in \mathcal{C}$ which is quite close to P^*.

After computing $\hat{P}$, we will then estimate κ, θ, γ by applying the inversion formulas (3.10) to $\hat{P}$.

3.4.4 Approximate maximum likelihood estimation of the asset price

Let $U_n = X_{nT}$ be the asset price data jointly observed with volatility data $V_n = Y_{nT}$, for $n = 0, 1, 2, \ldots, N$. We estimate the new drift parameter $T\mu$ by maximum likelihood given all the observations U_n and V_n. The Euler discretization of the asset price SDE (3.4) leads to the following approximation,

$$U_{n+1} \approx U_n + T\mu U_n + \sqrt{T}\sqrt{V_n}U_n\Delta Z_1(n), \quad n = 0, 1, 2, \ldots, N-1, \;\; U_0 = X_0 = x_0, \;\; (3.24)$$

where $\Delta Z_1(n) = Z_1(n+1) - Z_1(n)$, $X_0 = x_0$ is a fixed constant and $E[\Delta B_1(n)\Delta Z_1(n)] = \rho$. We want to determine μ such that the discrete approximation for the asset price best fits

the given data in a maximum likelihood sense where we fix the volatility observations V_n. The discretized SDE driving the asset price immediately yields,

$$\sqrt{T}\Delta Z_1(n) \approx \frac{\Delta U_n - \nu U_n}{U_n\sqrt{V_n}},$$

for $n = 0, 1, \ldots, N-1$. As above, maximization of the approximate log-likelihood is equivalent to minimizing in μ the sum of the $(\Delta Z_1(n))^2$, and this yields the estimator

$$\hat{\mu} = \frac{2}{NTd}\sum_{n=0}^{N-1}\frac{\Delta U_n}{U_n V_n}, \tag{3.25}$$

where the statistic d was defined in (3.12). It should be noted that we have computed the log-likelihood estimator for μ by maximizing the joint density for only the observations U_n keeping V_n fixed.

Under the so-called "no arbitrage" principle, the option price does not depend on the "rate of return" parameter μ. However, for completeness, we will also study the estimator $\hat{\mu}$ of μ, prove its consistency and evaluate its correlation with other parameter estimators

3.4.5 Estimation of the correlation between Brownian motions

After estimating the parameters of the volatility model and the drift μ, we now estimate the underlying Brownian motion increments $\Delta Z_1(n)$ and $\Delta B_1(n)$ by

$$\begin{aligned}\Delta Z_1(n) &\approx \frac{\Delta U_n - T\hat{\mu}U_n}{\sqrt{T}\sqrt{V_n}U_n},\\ \Delta B_1(n) &\approx \frac{\Delta V_n - (\hat{u} - \hat{v}V_n)}{\sqrt{2\hat{w}}\sqrt{V_n}}.\end{aligned} \tag{3.26}$$

The empirical estimator $\hat{\rho}$ of ρ is then naturally defined by,

$$\hat{\rho} = \frac{1}{N}\sum_{n=0}^{N-1}\left[\frac{\Delta U_n - \hat{\nu}U_n}{\sqrt{T}\sqrt{V_n}U_n}\right] \times \left[\frac{\Delta V_n - (\hat{u} - \hat{v}V_n)}{\sqrt{2\hat{w}}\sqrt{V_n}}\right]. \tag{3.27}$$

Therefore we have now obtained formulas for all the five estimators of the Heston model.

A complementary study was done to evaluate if there was a numerical improvement in the use of simultaneous maximum likelihood estimators based on the joint density of the discretized process over the decoupled estimation approach described in this chapter. We compared the accuracy of estimators and the CPU time for both the methods. We report on this in the Technical Annex.

Chapter 4

Parameter estimators : Asymptotic behavior

4.1 Asymptotics of the estimators when $T \to 0$ and $S = NT$ is fixed

Definition 4.1.1. *Let $Y \in \mathbb{R}^k$ be a random vector whose distribution has a density Φ_Θ depending on the parameter $\Theta \in \mathbb{R}^k$. Let $X = (X_1, X_2, \ldots, X_n)$ be a discrete time random process whose distribution depends on the parameter Θ. Let $\hat{\Theta}_n$ be an arbitrary estimator of Θ based on observation $X_1, X_2, \ldots, X_n$. Then the sequence $\hat{\Theta}_n$ is said to be a consistent sequence of estimators if $\hat{\Theta}_n \to \Theta$, in probability.*

The process Y_t satisfying the SDE (3.2) has a stationary Gamma distribution. However the transition density of the discrete process $V_n = Y_{nT}$ from the Euler approximation of the process Y_t does not satisfy Doeblin's condition for uniform ergodicity (see [28]) for fixed T. As a result for fixed T, the asymptotic properties of the estimator as the number of observations goes to infinity cannot be studied. We prove the consistency of the estimators by first letting $T \to 0$, and then letting $S \to \infty$.

4.2 Basic approximation results for generic SDEs

We first recall a few well known consequences of classical results for generic stochastic integrals (see P.A. Meyer's reference books [59]).

Theorem 4.2.1. *(see [59]). Fix a time interval* $[0, S]$. *Let* B_t *be a standard Brownian motion defined on a probability space* $(\Omega, \mathcal{F}_t, Pr)$, *endowed with a filtration* $\mathcal{F}_t$ *adapted to* B_t. *Let* $h(t)$, $j(t)$, $k(t)$ *be arbitrary almost surely continuous and non-anticipative random functions of time on the time interval* $[0, S]$.
Define the random continuous non-anticipative process $H(t)$, $J(t)$, $R(t)$ *by*

$$H(t) = \int_0^t h(t)dt, \quad J(t) = \int_0^t j(t)dB_t, \quad R(t) = H(t) + J(t).$$

Hence $R(t)$ *is the explicit solution starting at* $R(0) = 0$ *of the SDE*

$$dR(t) = h(t)dt + j(t)dB_t.$$

Note that $R = H + J$ *is the classical (see [59]) decomposition of* R *into a bounded variation component and a martingale component.*
For any integer N, *set* $T = S/N$ *and*

$$\Delta B(n) = B_{(n+1)T} - B_{nT}, \quad \Delta R(n) = R_{(n+1)T} - R_{nT}.$$

Define the discrete sums

$$J_N = \sum_{n=0}^{N-1} j(nT)\Delta B(n), \quad K_N = \sum_{n=0}^{N-1} k(nT)\Delta R(n), \quad M_N = \sum_{n=0}^{N-1} k(nT)(\Delta R(n))^2.$$

Then as $T = S/N \to 0$, the following limits hold for convergence in probability,

$$\begin{aligned}
&\lim_{T\to 0} J_N = \int_0^S j(t)dB_t, \\
&\lim_{T\to 0} K_N = \int_0^S k(t)dR(t) = \int_0^S k(t)h(t)dt + \int_0^S k(t)j(t)dB(t), \\
&\lim_{T\to 0} M_N = \int_0^S k(t)j(t)^2dt.
\end{aligned} \tag{4.1}$$

Proof. Note that the first basic result concerning $\lim_{T\to 0} J_N$ is a classical consequence of the definition of stochastic integrals with respect to Brownian motion (see [59]). Assume temporarily that $|h(t)|, |j(t)|, |k(t)|$ are bounded by a deterministic constants for $t \in [0, S]$. Then the results (4.1) are well known consequences of stronger classical theorems for generalized stochastic integrals (see for instance P.A. Meyer's books [59]).

In the more general case where there is no fixed deterministic bound for $|h(t)|, |j(t)|, |k(t)|$, select and fix an arbitrary large constant A, and let η be the (random) first time $t \in [0, S]$ at which one has $|h(t)| + |j(t)| + |k(t)| \geq A$, with the convention $\eta = S$ when there is no such $t \in [0, S]$.

Consider the set $\Omega(A)$ of all $\omega \in \Omega$ such that $\eta(\omega) = S$. The continuous random function $|h| + |j| + |k|$ has necessarily an almost surely finite random upper bound $\mathcal{A}$. Hence we have

$$\lim_{A\to+\infty} Pr[\Omega(A)] = Pr[\mathcal{A} \text{ is finite}] = 1.$$

Given any $\epsilon > 0$, we hence select and fix $A = A(\epsilon)$ large enough so that $Pr[\Omega(A)] > 1 - \epsilon$. On the set $\Omega(A) \subset \Omega$ the random functions and variables $\tilde{h}, \tilde{j}, \tilde{k}, \tilde{R}, \tilde{K}_N, \tilde{M}_N$ obviously are respectively identical to h, j, k, R, K_N, M_N. Define then the deterministically bounded continuous and non-anticipative random functions

$$\tilde{h}(t) = h(min(t, \eta)) \; ; \quad \tilde{j}(t) = j(min(t, \eta)) \; ; \quad \tilde{k}(t) = k(min(t, \eta)).$$

Replacing h, j, k by $\tilde{h}, \tilde{j}, \tilde{k}$ in the definitions of $R(t), K_N, M_N$ defines the corresponding $\tilde{R}(t), \tilde{K}_N, \tilde{M}_N$. As pointed out at the beginning of this proof, the limits in probability stated in (4.1) must hold for $\tilde{h}, \tilde{j}, \tilde{k}$ and $\tilde{R}, \tilde{K}_N, \tilde{M}_N$. Hence there is a number $T(A, \epsilon) > 0$ small enough, such that for each $T < T(A, \epsilon)$ one has $Pr[\Lambda(T)] > (1 - \epsilon)$, where the event $\Lambda(T)$ is the set of all $\omega \in \Omega$ such that

$$|\tilde{K}_N - \int_0^S \tilde{k}(t) d\tilde{R}(t)| \leq \epsilon, \quad |\tilde{M}_N - \int_0^S \tilde{k}(t)\tilde{j}(t)^2 dt| \leq \epsilon.$$

For $T < T(A(\epsilon), \epsilon)$, the event $\Omega_{\epsilon,T} = \Omega(A(\epsilon)) \cap \Lambda(T)$ has probability larger than $(1 - 2\epsilon)$, and for all $\omega \in \Omega_{\epsilon,T}$ the last inequalities can immediately be rewritten as

$$|K_N - \int_0^S k(t) dR(t)| \leq \epsilon, \quad |M_N - \int_0^S k(t) j(t)^2 dt| \leq \epsilon.$$

This achieves the proof. □

4.3 Summary of main notations

The processes X_t and Y_t driven by the two Heston SDEs (3.1)-(3.2) are associated to two standard Brownian motions Z_t, B_t, with instantaneous correlation ρ. Call $S = NT$ the available global observation time for X_t and Y_t, where T is the time interval between consecutive observations and N is the total number of observations. We will first study the asymptotics of our parameter estimators when S is fixed and $T \to 0$, so that $N \to \infty$.

Recall $V_n = Y_{nT}$ and $U_n = X_{nT}$ are the observed squared volatility and asset price respectively at time nT, for $n = 0, 1, \ldots, N$. The estimators for the parameters κ, θ, γ are algorithmically computed in terms of the statistics a, b, c, d, f, which were explicitly defined earlier in terms of the N observations $V_n = Y_{nT}$, by formulas (3.12).

4.4 Asymptotics of basic statistics a, b, c, d, f as $T \to 0$

We first compute the limits, for convergence in probability, of the statistics a, b, c, d, f as $T \to 0$ and $S = NT$ is fixed.

Proposition 4.4.1. *Consider the squared volatility Y_t solution of the Heston SDEs* (3.2)*, which verifies the autonomous SDE*

$$dY_t = \kappa(\theta - Y_t)dt + \gamma\sqrt{Y_t}dB_t,$$

where B_t is a standard Brownian motion. Consider N approximate observations $V_n = Y_{nT}$ of Y_t with $n = 0, \dots, N-1$. The estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ of κ, θ, γ are then computed by the algorithm of chapter 3 as deterministic functions of the statistics a, b, c, d, f, which are explicitly defined in terms of the V_n by the expressions (3.12)*.*
For fixed $S = NT$ we define the five random variables $\beta(S)$, $\chi(S)$, $\delta(S)$, $\phi(S)$, $\tau(S)$ by

$$\begin{aligned} \chi(S) =& \frac{2}{S}(Y_S - Y_0), \quad \delta(S) = \frac{1}{S}\int_0^S \frac{1}{Y_t}dt, \quad \phi(S) = \frac{1}{S}\int_0^S Y_t dt, \\ \tau(S) =& \frac{1}{S}\int_0^S \frac{1}{\sqrt{Y_t}}dB_t, \quad \beta(S) = 2\kappa - 2\kappa\theta\delta(S) - 2\gamma\tau(S). \end{aligned} \tag{4.2}$$

As $T \to 0$ and S is kept fixed, we have the following convergence in probability results

$$\begin{aligned} \lim_{T\to 0}(a/T) =& \gamma^2, \quad \lim_{T\to 0}(b/T) = \beta(S), \quad \lim_{T\to 0} c/T = \chi(S), \\ \lim_{T\to 0} d =& 2\delta(S), \quad \lim_{T\to 0} f = 2\phi(S). \end{aligned} \tag{4.3}$$

Proof. Define the random variables $\chi(S)$, $\delta(S)$, $\phi(S)$, $\tau(S)$ by the explicit expressions in (4.2) above. Formulas (3.12) give for statistic c the expression

$$c = \frac{2}{N}\sum_{n=0}^{N-1}\Delta V_n = \frac{2}{N}(V_N - V_0),$$

which obviously implies the following convergence in probability

$$\lim_{T\to 0} c/T = \frac{2}{S}(Y_S - Y_0) = \chi(S). \tag{4.4}$$

Statistics d and f have the expressions (see (3.12))

$$d = \frac{2}{S}\sum_{n=0}^{N-1}\frac{T}{V_n}, \quad f = \frac{2}{S}\sum_{n=0}^{N-1} TV_n.$$

The two sums above are the classical Riemann sums discretizing the Riemann integrals $\delta(S)$, $\phi(S)$, where the respective integrands are the almost surely continuous random functions $1/Y_t$ and Y_t. The definition of Riemann integrals for continuous functions then yields the following almost sure limits (and hence a fortiori the convergences in probability)

$$\lim_{T\to 0} d = 2\delta(S), \quad \lim_{T\to 0} f = 2\phi(S). \tag{4.5}$$

From the equation (3.12) we get

$$\begin{aligned} a/T =& \frac{1}{S}\sum_{n=0}^{N-1}\frac{(\Delta V_n)^2}{V_n}, \\ b/T =& \frac{-2}{S}\sum_{n=0}^{N-1}\frac{\Delta V_n}{V_n}. \end{aligned} \tag{4.6}$$

Since the squared volatility Y_t remains almost surely strictly positive when $Y(0) > 0$, the function $1/Y_t$ is almost surely continuous in t, and we can apply Theorem 4.2.1 with

$$R(t) = Y_t, \quad h(t) = \kappa(\theta - Y_t), \quad j(t) = \gamma\sqrt{Y_t}, \quad k(t) = 1/Y_t.$$

We then have with the notations of Theorem (4.2.1)

$$\frac{a}{T} = \frac{1}{S} M_N, \quad b/T = \frac{-2}{S} K_N.$$

Hence Theorem (4.2.1)now entails the following convergences in probability

$$\begin{aligned} \lim_{T\to 0} \frac{a}{T} = \frac{1}{S} \lim_{T\to 0} M_N =& \frac{1}{S} \int_0^S k(t) j(t)^2 dt = \gamma^2, \\ \lim_{T\to 0} \frac{b}{T} = \frac{-2}{S} \lim_{T\to 0} K_N =& \frac{-2}{S} [\int_0^S k(t) h(t) dt + \int_0^S k(t) j(t) dB(t)] \\ =& 2\kappa - 2\kappa\theta\delta(S) - 2\gamma\tau(S). \end{aligned} \tag{4.7}$$

□

4.5 Asymptotics of the random variables $\beta(S)$, $\chi(S)$, $\delta(S)$, $\phi(S)$, $\tau(S)$ for large global observation time S

Note that the explicitly defined random variables $\beta(S), \chi(S), \delta(S), \phi(S), \tau(S)$ involved in the limits as $T \to 0$ of our parameter estimators depend on the fixed global observation time S but do not depend on T, N any more.

Proposition 4.5.1. *Let Y_t be the squared volatility process solution of the Heston SDE (3.2) on $\mathbb{R}^+$. Then as the global observation time S tends to infinity, we have the following limits for convergence in probability*

$$\begin{aligned} \lim_{S\to\infty} \beta(S) =& -2\kappa\gamma^2/(2\kappa\theta - \gamma^2), \quad \lim_{S\to\infty} \chi(S) = 0, \\ \lim_{S\to\infty} \delta(S) =& 2\kappa/(2\kappa\theta - \gamma^2), \\ \lim_{S\to\infty} \phi(S) =& \theta, \quad \lim_{S\to\infty} \tau(S) = 0. \end{aligned} \tag{4.8}$$

Proof. Note first the L_2 - limit, implying a fortiori the convergence in probability

$$\lim_{S\to\infty} \chi(S) = \lim_{S\to\infty} Y_S/S = 0. \tag{4.9}$$

The probability distributions G_t of Y_t converges weakly to the stationary distribution G on $\mathbb{R}^+$, as $t \to \infty$ (see [32]), and G has a density function $g(y)$ given by

$$g(y) = \frac{\xi}{\Gamma(\nu)}(\xi y)^{\nu-1}e^{-\xi y}\mathbf{1}_{\{y>0\}}.$$

where

$$\xi = v/w = 2\kappa/\gamma^2, \quad \nu = u/w = 2\kappa\theta/\gamma^2, \quad \Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt.$$

One can uniformly control the tails of all the probabilities G_t on $\mathbb{R}^+$ near 0 and near $+\infty$, in order to show that

$$\lim_{S\to\infty} E[Y_S] = \int_0^\infty yg(y)dy, \quad \lim_{S\to\infty} E[\frac{1}{Y_S}] = \int_0^\infty \frac{1}{y}g(y)dy.$$

and the explicit form of $g(y)$ then easily provides the explicit expressions

$$\int_0^\infty yg(y)dy = \frac{\nu}{\xi} = \theta, \quad \int_0^\infty \frac{1}{y}g(y)dy = \frac{\xi}{(\nu-1)} = 2\kappa/(2\kappa\theta - \gamma^2).$$

Hence the random variable $\delta(S)$ verifies

$$\lim_{S\to\infty} E[\delta(S)] = lim_{S\to\infty}\frac{1}{S}\int_0^S E(\frac{1}{Y_t})dt = \lim_{S\to\infty} E[\frac{1}{Y_S}] = 2\kappa/(2\kappa\theta - \gamma^2).$$

and the ergodic theorem implies the following convergence in probability

$$\lim_{S\to\infty} \delta(S) = \lim_{S\to\infty}\frac{1}{S}\int_0^S \frac{1}{Y_t}dt = \lim_{S\to\infty} E[\frac{1}{Y_S}] = 2\kappa/(2\kappa\theta - \gamma^2).$$

By definition (4.2) of the stochastic integral $\tau(S)$, we have $E[\tau(S)^2] = \dfrac{E[\delta(S)]}{S}$. Since $E[\delta(S)]$ tends to a finite constant limit as $S \to \infty$, we conclude that $\lim_{S\to\infty} E[\tau(S)^2] = 0$, and hence a fortiori that $\lim_{S\to\infty} \tau(S) = 0$ for convergence in probability.

Again by the ergodic theorem, we have the following convergences in probability

$$\lim_{S\to\infty} \phi(S) = \lim_{S\to\infty} \frac{1}{S} \int_0^S Y_t dt = \lim_{S\to\infty} E[Y_S] = \theta.$$

The random variable $\beta(S)$ is defined in (4.2) as an explicit linear combination of $\delta(S)$ and $\tau(S)$ with fixed coefficients so that we immediately have the following convergence in probability

$$\lim_{S\to\infty} \beta(S) = -\frac{2\kappa\gamma^2}{2\kappa\theta - \gamma^2}.$$

□

4.6 Asymptotic behavior of parameter estimators for Heston SDEs

We have noted in chapter 3 that the maximization algorithm, which a priori may consider four cases, actually needs only to implement the Case 1 estimation formulas (3.17) as soon as the applicability conditions (3.16) are satisfied. For convenience we recall here these 3 conditions

$$2b + cd < bf + 2c < 0 \quad \text{and} \quad 2a(df - 4) - (b^2 f + 4bc + c^2 d) + 4(2c + bf) < 0.$$

As $T \to 0$ with S fixed, our preceding asymptotics results (4.3) have expressed the limits in probability of statistics $a/T, b/T, c/T, d, f$ in terms of five random explicitly defined random variables $\beta(S), \chi(S), \delta(S), \phi(S), \tau(S)$.

These asymptotic formulas (4.3) show that the three inequalities imposed on a, b, c, d, f

by the Case 1 feasibility conditions (3.16) are asymptotically equivalent to the following set of three asymptotic inequalities involving explicit polynomials in the random variables $[\beta(S), \chi(S), \delta(S), \phi(S), \tau(S)]$, namely

$$\begin{aligned} Q_1(S) =& \beta(S)(1-\phi(S)) + \chi(S)(\delta(S)-1) < 0, \\ Q_2(S) =& \beta(S)\phi(S) + \chi(S) < 0, \\ Q_3(S) =& \gamma^2(\delta(S)\phi(S)-1) + \chi(S) + \beta(S)\phi(S) < 0. \end{aligned} \tag{4.10}$$

Proposition 4.6.1. *Consider the three random variables $Q_1(S), Q_2(S), Q_3(S)$ defined by* (4.10)*, which depend only on S. We have the following convergences in probability*

$$\lim_{S\to+\infty} Q_j(S) < 0 \quad \textit{for } j = 1, 2, 3$$

and hence for large global duration time S the asymptotic applicability conditions (4.10) *for Case 1 are satisfied with probability tending to 1 as $S \to \infty$.*

Proof. The convergence in probability of $\beta(S), \chi(S), \delta(S), \phi(S), \tau(S)$ to finite limits as $S \to \infty$ naturally extends to the corresponding convergence in probability for any polynomial in those five random variables. Hence to compute the limit in probability of $Q_j(S)$ as $S \to \infty$, we simply replace in the polynomial $Q_j(S)$, all the variables $\beta(S), \chi(S), \delta(S), \phi(S)$ by their explicit constant limits, which were computed above in (4.8). This immediately implies

$$\begin{aligned} \lim_{S\to\infty} Q_1(S) =& -2(1-\theta)\kappa\gamma^2/(2\kappa\theta-\gamma^2) < 0, \\ \lim_{S\to\infty} Q_2(S) =& -2\theta\kappa\gamma^2/(2\kappa\theta-\gamma^2) < 0, \\ \lim_{S\to\infty} Q_3(S) =& -\gamma^2 < 0. \end{aligned}$$

□

In the background probability space Ω on which all our processes and random variables

are defined, denote by Ω_S the following subset

$$\Omega_S = \{\omega \in \Omega \mid Q_j(S) < 0 \quad \text{for } \ j = 1, 2, 3\}.$$

In view of the last proposition, for any fixed small $\epsilon > 0$, we can find a global observation time $S(\epsilon)$ such that for each fixed $S > S(\epsilon)$, we will have

$$Pr(\Omega_S) = Pr(Q_j(S) < 0 \quad \text{for } \ j = 1, 2, 3) > 1 - \epsilon.$$

From now on we fix $S > S(\epsilon)$ and we will now focus only on the set of random trajectories in Ω_S .

We now show that as $T \to 0$, the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ computed by the Case 1 formula (3.17) have limits in probability. The estimation formula (3.17) combined with the inversion formula (3.8) linking u, v, w and κ, θ, γ show that

$$\hat{\kappa} = -\frac{2b + cd}{T(df - 4)}, \quad \hat{\theta} = \frac{bf + 2c}{2b + cd}, \quad \hat{\gamma} = \sqrt{[\frac{a}{T} - \frac{b^2 f + 4bc + c^2 d}{2T(df - 4)}]}.$$

Combining these formulas with the limits in probability of $a/T, b/T, c/T, d, f$ obtained above as $T \to 0$ with $S = NT$ fixed, yields for all random trajectories in Ω_S, the following limits for convergence in probability

$$\lim_{T \to 0} \hat{\kappa} = -\frac{\beta + \chi\delta}{2(\delta\phi - 1)}, \quad \lim_{T \to 0} \hat{\theta} = \frac{\beta\phi + \chi}{\beta + \chi\delta}, \quad \lim_{T \to 0} \hat{\gamma} = \gamma, \tag{4.11}$$

where the argument S has been omitted for brevity. In particular we see that on the subset Ω_S of the probability space Ω, the estimator $\hat{\gamma}$ of γ is asymptotically unbiased as $T \to 0$ with S fixed.

The other parameter estimators $\hat{\kappa}$ and $\hat{\theta}$ constructed here still have an asymptotic bias as T tends to zero with S fixed. Their limits in probability as $T \to 0$ with S fixed, defined by (4.11) above for trajectories in the (large) set Ω_S, can be considered as new estimators $\hat{\kappa}_{cont}(S)$, $\hat{\theta}_{cont}(S)$ for all the random trajectories belonging to the (large) set Ω_S. Hence we define these new estimators for all trajectories in Ω by the explicit formulas

$$\hat{\kappa}_{cont}(S) = -\frac{\beta + \chi\delta}{2(\delta\phi - 1)}, \quad \hat{\theta}_{cont}(S) = \frac{\beta\phi + \chi}{\beta + \chi\delta}. \tag{4.12}$$

We note that these "continuous time " estimators are essentially the maximum likelihood estimators when the whole process trajectory $Y_t, 0 \leq t \leq S$ is known. We can now compute the following limits in probability

$$\begin{aligned}
\lim_{S\to\infty}[\beta + \chi\delta] &= -\frac{2\kappa\gamma^2}{2\kappa\theta - \gamma^2} < 0, \\
\lim_{S\to\infty}[\beta\phi + \chi] &= -\theta\frac{2\kappa\gamma^2}{2\kappa\theta - \gamma^2} < 0, \\
\lim_{S\to\infty}[\delta\phi - 1] &= \frac{2\kappa\theta}{2\kappa\theta - \gamma^2} - 1 = \frac{\gamma^2}{2\kappa\theta - \gamma^2}
\end{aligned} \tag{4.13}$$

and this immediately implies the convergences in probability

$$\lim_{S\to\infty} \hat{\kappa}_{cont}(S) = \kappa, \quad \lim_{S\to\infty} \hat{\theta}_{cont}(S) = \theta. \tag{4.14}$$

Therefore, we have proved that on the large set Ω_S, the estimators $\hat{\kappa}, \hat{\theta}$ and $\hat{\gamma}$ converge in probability to the true parameters as $T \to 0$ and $S \to \infty$.

The estimators $\hat{\mu}$ and $\hat{\rho}$ can similarly be shown to converge in probability to their true parameter values. The asymptotic limit of $\hat{\mu}$ follows from the corresponding limit of the statistic d and an application of Theorem 4.2.1 and Proposition (4.8). The asymptotic limit of ρ follows as a result of the corresponding limits of all the other estimators together with an application of Theorem 4.2.1 and Proposition (4.8).

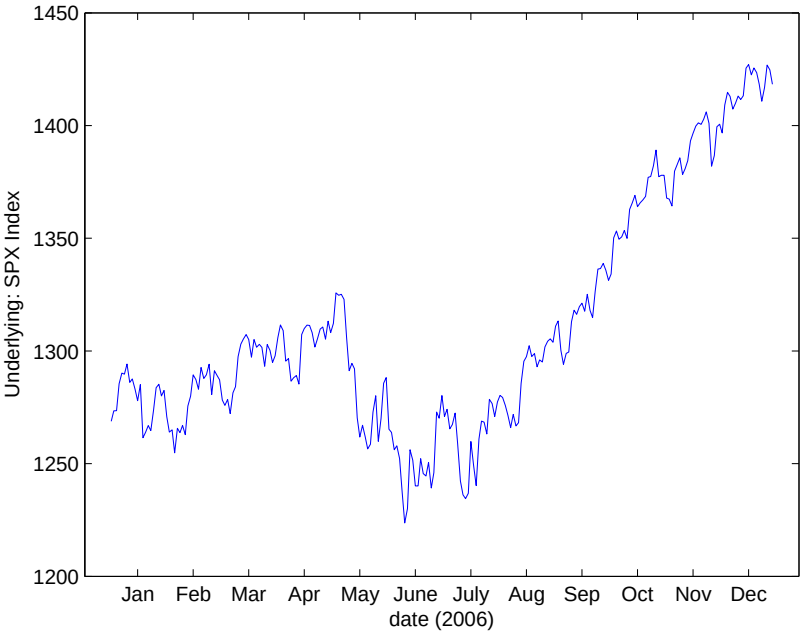
Chapter 5

Parameter estimation : Numerical tests on simulated and real data

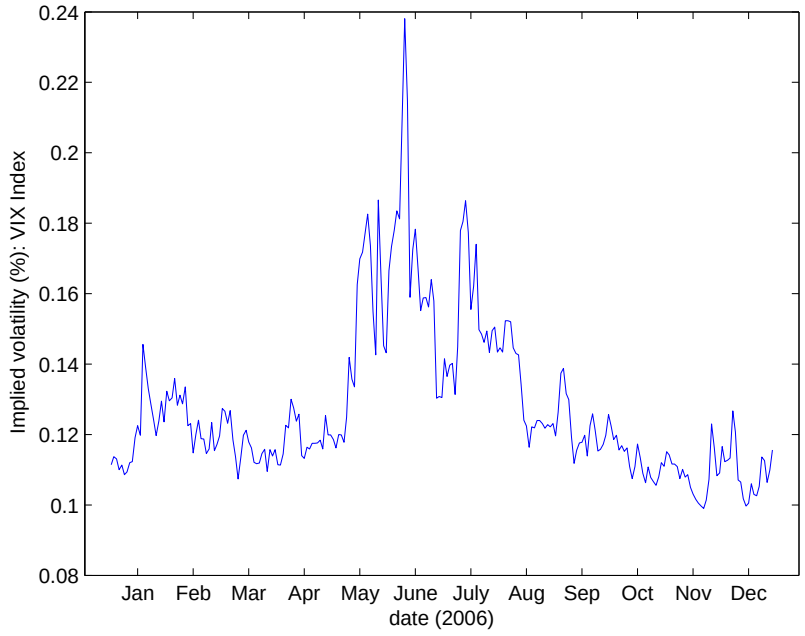
5.1 Fitting the Heston model to actual stockmarket data

We present numerical Heston model fitting for the daily S&P 500 index together with its volatility approximation, namely the VIX index for the period Jan 03 2006 to December 29 2006. This gives us a set of $N = 252$ observed data points. The parameter estimation and error computations can be similarly carried out for the Heston model fitted to intra-day S&P 500 data.

The methodology for the computation of the VIX index is described in the VIX white pages [72]. The Chicago Board Options Exchange computes and maintains a record of historical VIX index values. We take daily closing data for both indices in order to simultaneously observe the two data series [18]. The evolution of SPX and VIX over this period is displayed in Fig. 5.1(a) and Fig. 5.1(b). According to what would be a fairly common practice in concrete modeling of stock data by SDEs, we choose $T = 1/252$ as the time separating two consecutive observations.



(a) The SPX (S&P 500) index is the asset price value $\{X_t\}$



(b) The VIX index is an approximation of volatility of S&P 500 index.

Figure 5.1: Asset price and volatility

Using the estimation algorithms described in chapter, we obtain the following estimates for the parameters u, v, w

$$\hat{u} = .0011, \quad \hat{v} = .0664, \quad \hat{w} = 1.59710^{-4}.$$

By relation (3.8) we then compute the following estimates for the parameters of the Heston SDEs (3.1)-(3.2),

$$\hat{\kappa} = 16.6, \quad \hat{\theta} = .017, \quad \hat{\gamma} = .28, \quad \hat{\rho} = -.54, \quad \hat{\mu} = .102. \tag{5.1}$$

The negative value of ρ is linked to the fact that the asset price and squared volatility are negatively correlated. This is termed as the 'skew' or 'leverage' effect in financial jargon. When asset prices are high the squared volatility is low and when asset prices are low the squared volatility tends to be high.
The parameter estimation algorithm outlined above identifies Case 1 as the applicable case for this set of S&P data. The estimation algorithm is computationally very fast, because all our estimators are available in closed forms. The CPU-time to estimate the parameters from 252 data points is .002 seconds on a standard PC.

We will consider this estimated Heston model as a benchmark example to study numerically the small sample properties of our estimators and later to verify numerically our results on the asymptotic properties of the estimators. We will refer to the Heston model numerically parametrized by (5.1) as the 2006 S&P 500 model. Before presenting the results we describe the simulation method we will use to generate joint random trajectories for the asset price and squared volatility.

5.2 Simulation of the Heston SDEs

We fix a "global observation time" $S = NT$ for the joint processes (X_t, Y_t), where N is the number of discrete observations, and T the time interval between successive observations of the squared volatility Y_t.

To study small sample behavior and empirical asymptotics of our parameter estimators, we have implemented and analyzed Monte-Carlo simulations of the discretized SDEs driving jointly the asset price and squared volatility (see (3.24), (3.11)). Our objective is to first simulate trajectories which indeed are good approximations of the true continuous squared-volatility trajectory Y_t. We will denote by W_n this simulated discrete random trajectory emulating the squared volatility process Y_t, where W_n approximates $Y_{n\delta}$. The choice of the simulation step δ depends on the level of simulation accuracy we desire. In order to obtain a good approximation to the true process, δ has to be chosen sufficiently small. A discussion on the results for the convergence of W to Y is presented below.

The simulation of W_n is initiated by $W_0 = V_0 = y_o > 0$, and recursively generated by the equation

$$\Delta W_n = W_{n+1} - W_n = \delta\kappa(\theta - W_n) + \gamma\sqrt{W_n}\Delta B(n) \quad \text{for} \;\; n < STOP. \tag{5.2}$$

where $\Delta B_n = (B((n+1)\delta) - B(n))$ is a sequence of independent Gaussian random variables with mean zero and variance δ, and $STOP$ is the random first time when $W_n < 0$. We fix a small simulation step δ, and an associated integer m such that $T = m\delta$. After simulating the discrete random trajectory W_n by (5.2), we generate an associated random sample of N "virtual" squared volatility observations $V_k = Y_{kT}$, by setting $V_0 = W_0 = Y_0 = y_0 > 0$

and

$$V_k = Y_{kT} = W_{km\delta}.$$

For T small, the approximation (3.11) holds for V_k. As mentioned earlier for the study of S&P data the choice of T is the fairly small value $T = 1/252$.

It can be shown that for $S = NT = mN\delta$ fixed, the probability $Pr(STOP > Nm)$ tends to 1 as δ tends to zero. The recurrence relation (5.2) defines a time homogeneous discrete parameter Markov chain, where the conditional transition density of W_{n+1} given $W_n = x$ is Gaussian with mean $m(x) = x + u - vx$ and variance $var(x) = 2wx$.

Direct Monte Carlo simulations of the Euler discretized dynamics W_n without the restriction $n < STOP$ would ultimately generate negative values of W_n even when the parameters satisfy the imposed constraint $u > w$. So when simulating a discrete trajectory of fixed length mN, we naturally stop the simulated dynamics whenever W_n becomes negative for some $n \leq mN$. The corresponding truncated trajectory is then eliminated, and to replace it, we generate a new trajectory.

The discretized asset price process $\tilde{U}_n$ approximating $U_n = X_{n\delta}$ is then generated by the recursive relation

$$\tilde{U}_{n+1} = \tilde{U}_n + \mu\delta\tilde{U}_n + \sqrt{V_n}\tilde{U}_n(\rho B_n + \sqrt{1-\rho^2}K_n), \quad U_0 = X_0 = x_0,$$

where the K_n are independent Gaussian random variables with mean zero and variance δ, and are independent of all the B_m.

For $0 \leq t \leq S$, define non-anticipative left continuous processes (see [59] for definition) W_t and $\tilde{U}_t$ by,

$$\tilde{U}_t = \tilde{U}_n \quad \text{for} \;\; n\delta \leq t < (n+1)\delta; \quad W_t = W_n \quad \text{for} \;\; n\delta \leq t < (n+1)\delta.$$

In [45], the authors prove that the discrete process $\{W_n\}$ has finite second moments. For $S = mN\delta$ fixed, they also show that as $\delta \to 0$, W_t converges in L_2 to the squared volatility Y_t, with uniform L_2- speed of convergence over all $t \in [0, S]$.
This result can be extended to show also that as $\delta \to 0$ with S fixed, the discretized approximating process $\tilde{U}_t$ converges in L_2 to X_t for each $t \in [0, S]$.

5.3 Small sample bias and variances of parameter estimators

We study the small sample bias and errors in the estimators of the Heston model, denoted here "HSP", parametrized by the numerical values (5.1), values which were deduced above from our sample of 252 joint observations of the 2006 S&P daily data.
We first fix a very small simulation time step $\delta = 1/25,200$ and the time interval $T = 100\delta = 1/252$ between virtual observations of the squared volatility. We then simulate 5000 paths of the Heston model HSP using the small discretization time step δ. We then sub-sample each one of these 5000 trajectories, at time intervals $T, 2T, \ldots, NT$ to generate N "virtual" observations $V_k = Y_{km\delta}$. For each one of these virtual V-trajectories of length N, we compute an estimate of the parameters, thereby generating a random sample of size 5000 for each parameter estimator. For each parameter, we then compute the empirical mean and standard deviation of the associated estimator over these 5000 trajectories. The *Bias* of an estimator is classically the expected value of the difference between the estimator and the true parameter value, where the expected value is taken with respect to the true underlying probability distribution Pr. The estimated bias of a parameter estimator is then

the difference between its empirical mean and its target numerical value given by (5.1).

The *estimation error* ERR of a parameter estimator is here defined as the L_2 norm of the difference between the estimator and the corresponding true parameter value. One always has the elementary relation

$$(ERR)^2 = Bias^2 + Variance. \tag{5.3}$$

In the numerical examples below, to evaluate the estimation error ERR, we replace bias and variance by their empirical estimates computed from the sample of 5000 estimator values generated above.

In Table 5.1, we fix $N = 252$, $T = 1/252$, and hence $S = NT = 1$, and we present, for the 4 parameter estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}, \hat{\rho}$, accurate empirical estimates of their mean, relative bias, relative standard deviation and relative error. Here the term "relative" stands for "relative to the mean" of the estimator. Clearly this natural notion is only useful when the mean of the estimator is not too small. Note also that the relative bias is here always presented by its modulus, since its sign is not relevant. To simulate 5000 trajectories of $N = 252$ observations generated using $\delta = 1/25200$ and $T = 100\delta$, the CPU-time is 30 seconds on a standard PC. To generate the corresponding sample of 5000 parameter estimators values, the CPU-time is 5 seconds on a standard PC. We present separately the estimation error for the estimator $\hat{\mu}$.

For the Heston model HSP, we also present, in Table 5.1, results for larger values of the number of observations N, namely $N = 504,\ S = NT = 2$; and $N = 1008,\ S = NT = 4$. Pragmatically, an SDE model with constant parameters for the S&P data cannot remain accurate over periods of several years. Therefore values of N larger than 300 for daily stock and volatility observations are definitely unrealistic. We nevertheless present results for

$N > 300$ to illustrate the decrease in estimation errors.

	N=252				N=504				N=1008			
	Mean	|Bias|/ Mean	Std/ Mean	Rel. Error	Mean	|Bias|/ Mean	Std/ Mean	Rel. Error	Mean	|Bias|/ Mean	Std/ Mean	Rel. Error
$\hat{\kappa}$	20.1	18%	33%	38%	18.1	8%	24%	25%	17.3	4%	20%	20%
$\hat{\theta}$	.017	.3%	13%	13%	.017	.2%	9%	9%	.017	.2%	7%	7%
$\hat{\gamma}$	.273	4%	5%	5%	.273	3%	3%	4%	.274	3%	3%	4%
$\hat{\rho}$	-.543	.2%	11%	11%	-.544	.02%	8%	8%	-.545	.1%	6%	6%

Table 5.1: Small sample relative accuracy of parameter estimation for different values of N. True parameter values : $\kappa = 16.6$, $\theta = .017$, $\gamma = .2826$, $\rho = -.5441$, $\mu = .1017$, $T = 1/252 = .004$.

The value of μ for the Heston model HSP is quite small, so that the mean of $\hat{\mu}$ is then also quite small, and the relative estimation error for $\hat{\mu}$ is practically meaningless, particularly for moderate values of N. Therefore we will only consider absolute bias, standard deviation, and L_2 error for $\hat{\mu}$, which we present along with the absolute bias, standard deviation and L_2 error of the other parameters in Table 5.2. We observe that as N increases the absolute errors in all parameters indeed decrease.

We point out that in chapter 7 when we study the impact of estimation errors on option price we will only need the absolute errors in the parameter estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}, \hat{\rho}$. We will not need to use the estimation errors of the estimator $\hat{\mu}$ because option prices do not depend on μ.

	N=252				N=504				N=1008			
	Mean	|Bias|	Std	Error	Mean	|Bias|	Std	Error	Mean	|Bias|	Std	Error
$\hat{\kappa}$	20.1	3.54	6.8	7.66	18.1	1.45	4.4	4.60	17.3	.72	3.4	3.5
$\hat{\theta}$	.017	.0001	.0022	.0022	.017	$3.\ 10^{-5}$	.0016	.0016	.017	3.10^{-5}	.0013	.0013
$\hat{\gamma}$	.273	.010	.012	.016	.273	.009	.001	.013	.274	.009	.007	.011
$\hat{\rho}$	-.543	.001	.059	.059	-.544	.0001	.041	.041	-.545	.0008	.034	.034
$\hat{\mu}$	.091	.0105	.122	.1227	.096	.006	.085	.085	.097	.004	.070	.070

Table 5.2: Small sample absolute accuracy of parameter estimation for different values of N. True parameter values : $\kappa = 16.6,\ \ \theta = .017,\ \ \gamma = .2826,\ \ \rho = -.5441,$ $\mu = .1017,\ \ T = 1/252 = .004.$

5.4 Numerical results on consistency of parameter estimators

We show by extensive numerical simulations that the asymptotic consistency results proved in chapter 4 are validated numerically for the Heston model parameters κ, θ and γ. We study the asymptotic estimation errors of our parameter estimators for a family of $8 = 2 \times 2 \times 2$ arbitrary Heston models.

For more pragmatic pertinence we choose the parameter vectors of our 8 benchmark Heston models in a moderately large neighborhood of the Heston model HSP estimated from our 252 joint daily observations of the 2006 S&P 500 and VIX data (5.1). The

parameter values we have selected for our 8 benchmark Heston models are listed below.

κ	θ	γ
16.6	.017	.28
16.6	.017	.10
16.6	.50	.28
16.6	.50	.10
25	.017	.28
25	.017	.10
25	.50	.28
25	.50	.10

(5.4)

For each one of these 8 benchmark models, we fix the parameters $\mu = .01$ and $\rho = -.54$, as in (5.1).

We simulate 1000 random trajectories for each one of our 8 benchmark Heston models, using the method described in section 5.2, with a very small simulation step $\delta = 1/10000$. The T value will then have to be larger than $10 \times \delta$ to maintain good simulation accuracy by (3.11). To decide on a reasonable ratio T/δ for accurate simulations we evaluate the Gaussianity of the random variables,

$$Q_n^{T,S} = \frac{\Delta V_n^{T,S} - T\kappa(\theta - V_n^{T,S})}{\gamma\sqrt{V_n^{T,S}}\sqrt{T}},$$

where $\{V_n^{T,S}\}$ denotes the observations up to time S sampled at intervals of length $T = m\delta$. Recall, that with large probability the Case 1 algorithm is the only one we need to apply to estimate the parameters κ, θ and γ. We observe numerically from the 1000 simulated trajectories of each one our 8 benchmark models that indeed the algorithm Case 1 is applicable with probability one for various pairs of reasonable values of T and S.

We have shown in the previous section that for a fixed S, as $T \to 0$, the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ converge in probability to estimators $\kappa_{cont},\ \theta_{cont},\ \gamma_{cont}$, corresponding to the continuous process Y_t. If we then let $S \to \infty$, the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ indeed converge in probability to the true parameters, i.e.,

$$Pr(|\hat{\kappa}_{cont}(S) - \kappa| < tol) \to 1, \quad Pr(|\hat{\theta}_{cont}(S) - \theta| < tol) \to 1,$$

$$Pr(|\hat{\gamma}_{cont}(S) - \gamma| < tol) \to 1.$$

as $S \to \infty$. We use the following notations to denote the probabilities of the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ being within a prescribed tolerance of their ideal limit values κ, θ, γ respectively.

$$Pr_\kappa(S,T) = Pr(|\hat{\kappa} - \kappa| < tol_\kappa), \quad Pr_\theta(S,T) = Pr(|\hat{\theta} - \theta| < tol_\theta),$$

$$Pr_\gamma(S,T) = Pr(|\hat{\gamma} - \gamma| < tol_\gamma).$$

Clearly these probabilities depend on the choice of the prescribed tolerances but for brevity our notations deliberately omit to keep track of the tolerances. We will explicitly specify the value of each of tol_κ, tol_θ and tol_γ for our results. Note that our choice of tolerance level for each parameter should naturally depend on the size of the parameter. We illustrate with numerical examples the theoretical results for each of our 8 benchmark models. We first study numerically the convergence in probability of $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ to κ, θ, γ. Then we present a comparative study of this convergence in probability for each one of our 8 benchmark models.

5.5 Asymptotic study of model estimated from stockmarket data

We present first numerical asymptotic results for the Heston model HSP associated to the 2006 S&P 500 data, with parameter values given by the first row of (5.4). For the errors of estimation of $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$, we fix "acceptable tolerance levels" as follows

$$tol_\kappa = .83, \quad tol_\theta = 8.10^{-4}, \quad tol_\gamma = .014.$$

We fix a very small discretization time step $\delta = 1/10,000$ and a small value $T = 1/1000 = 10\delta$. We then simulate $N = S/T$ virtual observations V_n of the squared volatility at time intervals nT, $(n = 1, 2, \ldots, N)$. We study the asymptotic behavior of $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ as $S = NT$, and hence N, increase.

Fig. 5.2, 5.3, and 5.4 present the values of $Pr_\kappa,\ Pr_\theta,\ Pr_\gamma$ respectively as $S = NT$ increases for fixed $T = 1/1000$. We compute the empirical probabilities $Pr_\kappa,\ Pr_\theta,\ Pr_\gamma$ as S increases from 0.5 to 20 by increments of 1.5. The corresponding values of the number of observations N increase from $N = 500$ to $N = 20,000$ by increments of $1,500$.

For each S, we have indeed verified that Case 1 is applicable with empirical probability equal to 1.

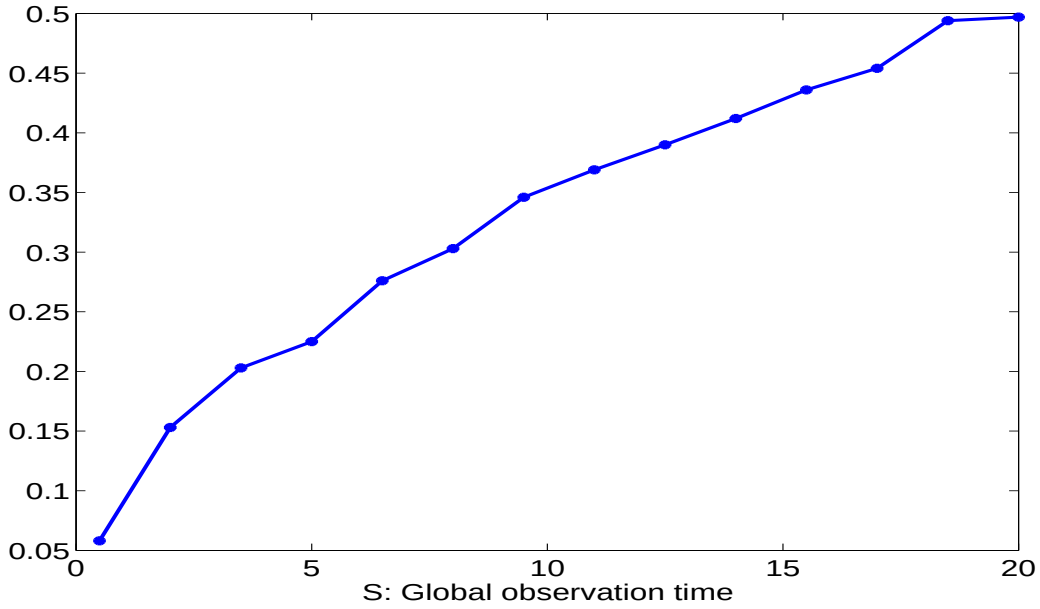


Figure 5.2: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$, fixed $T = 1/1000$. Slow increase to 1/2 for Pr_κ, as $N = S/T$ increases from 500 to 20,000.

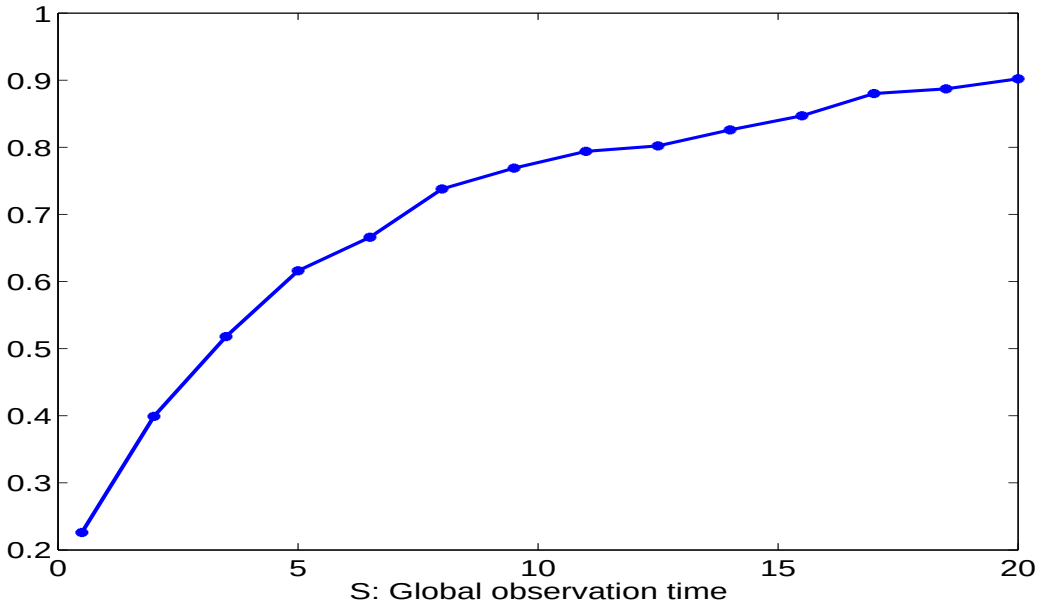


Figure 5.3: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$, fixed $T = 1/1000$. Convergence towards 1 of Pr_θ as $N = S/T$ increases from 500 to 20,000.

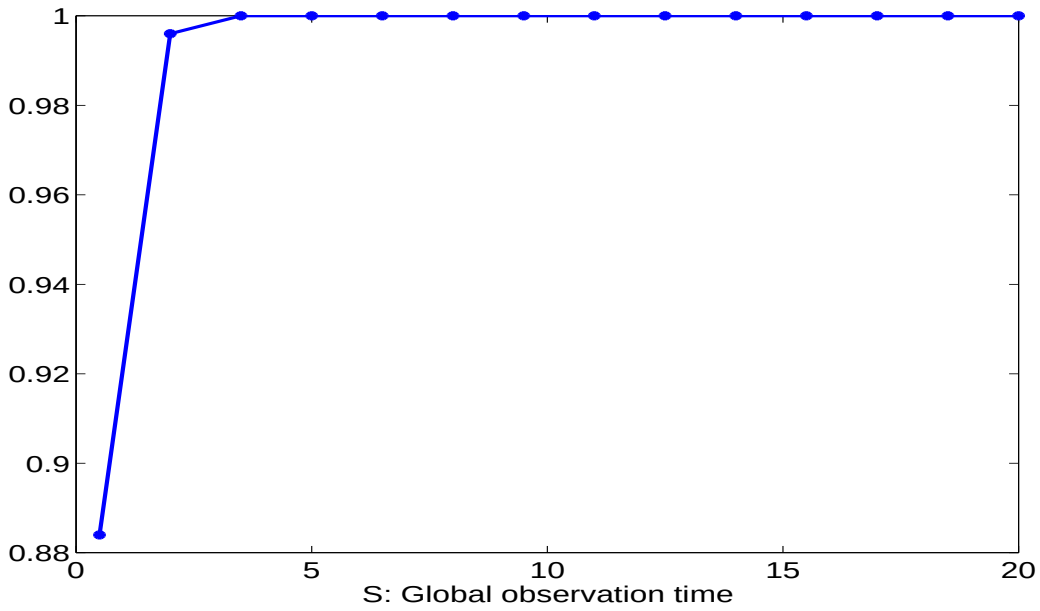


Figure 5.4: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$, fixed $T = 1/1000$. Fast Convergence to 1 of Pr_γ as $N = S/T$ increases from 500 to 20,000.

We observe that for fixed $T = 1/1000$ the probabilities $Pr_\theta(S,T)$ and $Pr_\gamma(S,T)$ indeed converge to one as N increases from $N = 500$ to $N = 20,000$, while for $Pr_\kappa(S,T)$ to converge to one, much larger number of observations are needed for adequate convergence. We find indeed that about $N = 60,000$ observations are needed for adequate convergence of $\hat{\kappa} - \kappa$ within a tolerance of 0.83. We observe on our benchmark models that $\hat{\kappa}$ converges faster for larger values of θ. This is illustrated in Fig. 5.5 where we fix $\kappa = 16.6, \gamma = .28$ and vary θ. In Fig. 5.5 the data are simulated at $T = 1/1000$, with S increasing from 0.5 to 10, corresponding to N increasing from 500 to $10,000$

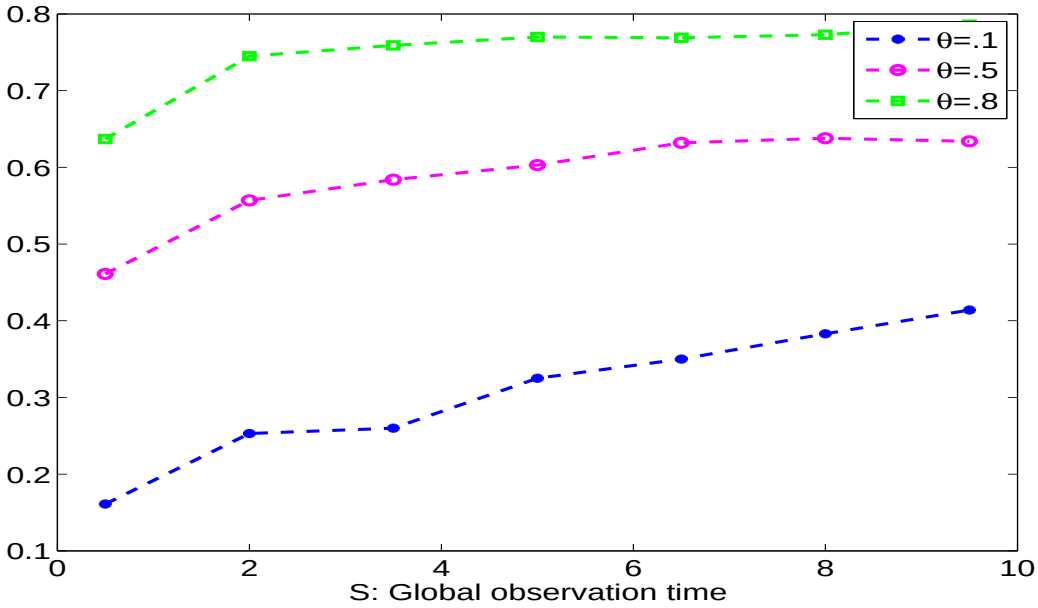


Figure 5.5: The ordinate represents the value of Pr_{κ}. We observe higher values of Pr_{κ} for larger values of θ.

In the next set of results we study the asymptotic behavior of the estimators when the global observation time $S = NT$ remains fixed, and the sampling time interval T tends to 0. We compute the values of the parameter estimators for sampling times varying from $T = 10\delta = .001$ to $T = 960\delta = .096$ with increments of $T = 50\delta = .005$.

We illustrate the results for the convergence to one of Pr_κ, Pr_θ, Pr_γ respectively in Fig. 5.6, Fig. 5.7, and Fig. 5.8. In each figure we illustrate the probabilities corresponding to $S = 10,\ 20,\ 30$. The probabilities Pr_κ, Pr_θ, Pr_γ are estimated over 1000 simulated trajectories.

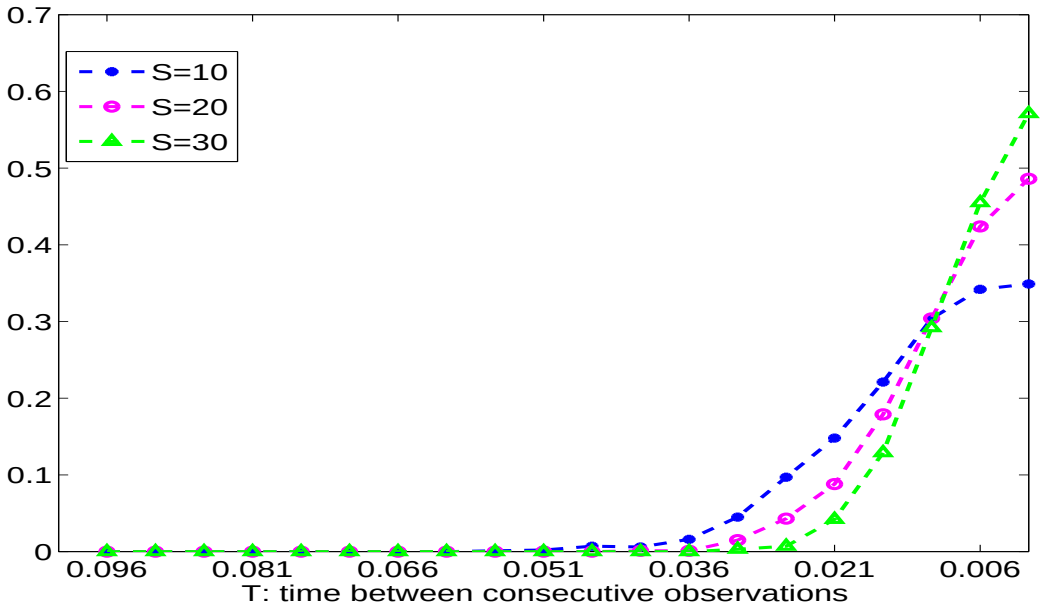


Figure 5.6: Slow "Convergence to 1" of Pr_κ as T decreases from .096 to .001. Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$.

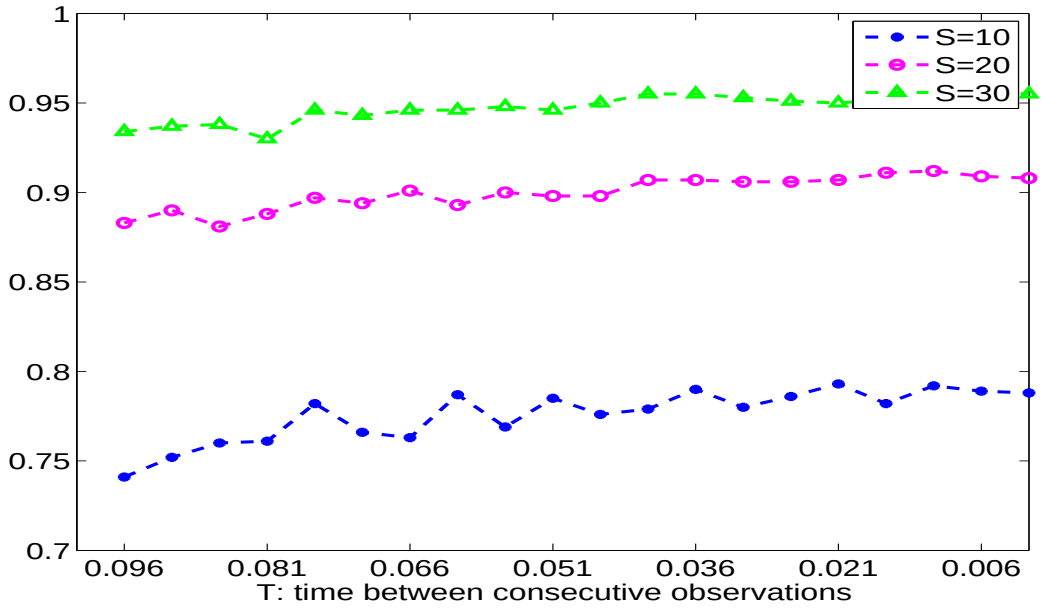


Figure 5.7: Convergence to 1 of Pr_θ as T decreases from .096 to .001. Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$.

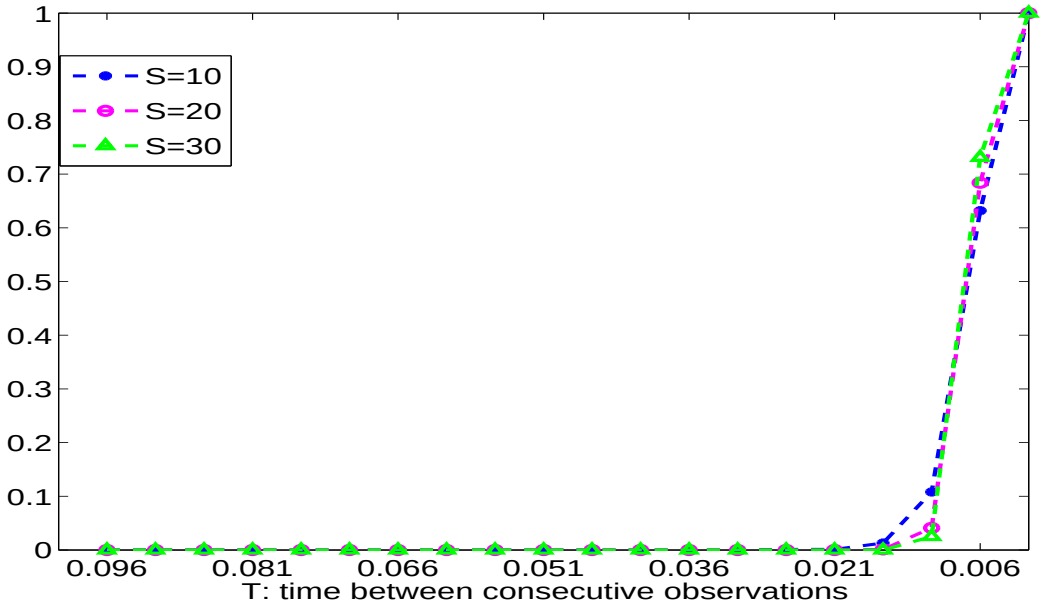


Figure 5.8: Convergence to 1 of Pr_γ as T decreases from .096 to .001. Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$.

We see that the estimator $\hat{\gamma}$ converges in probability to γ at the tolerance level tol_γ for each fixed S as T becomes smaller. This validates our theoretical convergence result (4.12) for $\hat{\gamma}$.

For $\hat{\kappa}$ and $\hat{\theta}$ we know from (4.12) that for each fixed S the estimators $\hat{\kappa}, \hat{\theta}$ converge to the continuous time estimators $\hat{\kappa}_{cont}, \hat{\theta}_{cont}$, which are random variables with non zero dispersions. The numerical results show that for fixed large enough values of S such as $S = 30$ corresponding to a fixed large enough $N = 30,000$, the estimators $\hat{\kappa}$ and $\hat{\theta}$ converge in probability to the true parameter values at the tolerance levels tol_κ and tol_θ respectively. However, for $\hat{\kappa}$ we need to simulate observations for much longer observation times S in order to achieve a probabilistic accuracy comparable to the accuracy of $\hat{\theta}$.

We observe that for $\hat{\kappa}$ and $\hat{\gamma}$ the accuracy of the estimators quickly deteriorates as the sampling time interval T increases. On the other hand for the estimator $\hat{\theta}$, the accuracy is maintained even at very high sampling steps.

5.6 Comparative study of eight benchmark Heston models

Our second set of results is a comparative study of the 8 different models associated to the parameter values given in (5.4). In order to carry out a comparative study we fix the tolerance level to be 5% of the true parameter value for each of the three parameters. We choose the tolerance error as a fixed percentage of the parameter values only to facilitate numerical comparison of results. The numerical results in this section show the convergence of the probabilistic accuracies $Pr_\kappa, Pr_\theta, Pr_\gamma$ to one for fixed $T = 1/1000$ when S increases from 0.5 to 30 by increments of 1.5. This corresponds to values of N increasing from 500 to $30,000$ by increments of $1,500$.

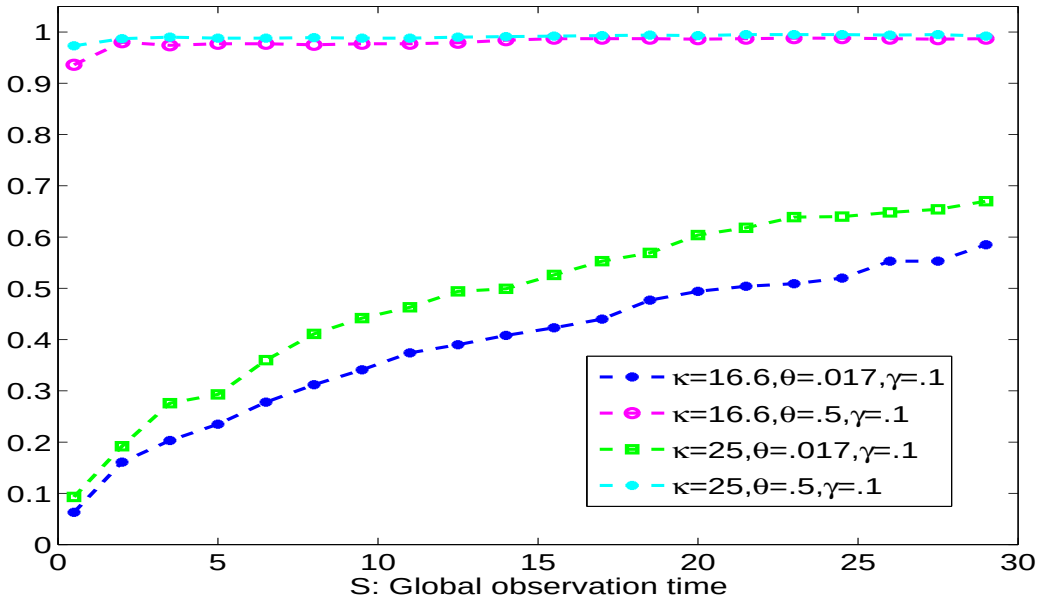


Figure 5.9: Convergence to 1 of Pr_κ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1,000$ and $\gamma = .1$.

We first fix $\gamma = .1$. We then perform the preceding simulations and studies for the 4 benchmark models for which $\gamma = 0.1$.

Fig. 5.9 and Fig. 5.10 illustrate the asymptotic behavior of the estimators $\hat{\kappa}$ and $\hat{\theta}$. Each figure displays 4 curves (one per benchmark Heston model) plotting the probability that a specific parameter estimator falls within the preassigned tolerance interval centered at the true parameter value. The legends describe the parameter values associated to each curve. We have not displayed the convergence of Pr_γ to 1 because for each one of our 4 benchmark models, we have $Pr_\gamma = 1$ as soon as $S \geq 15$, which corresponds to $N \geq 15,000$.

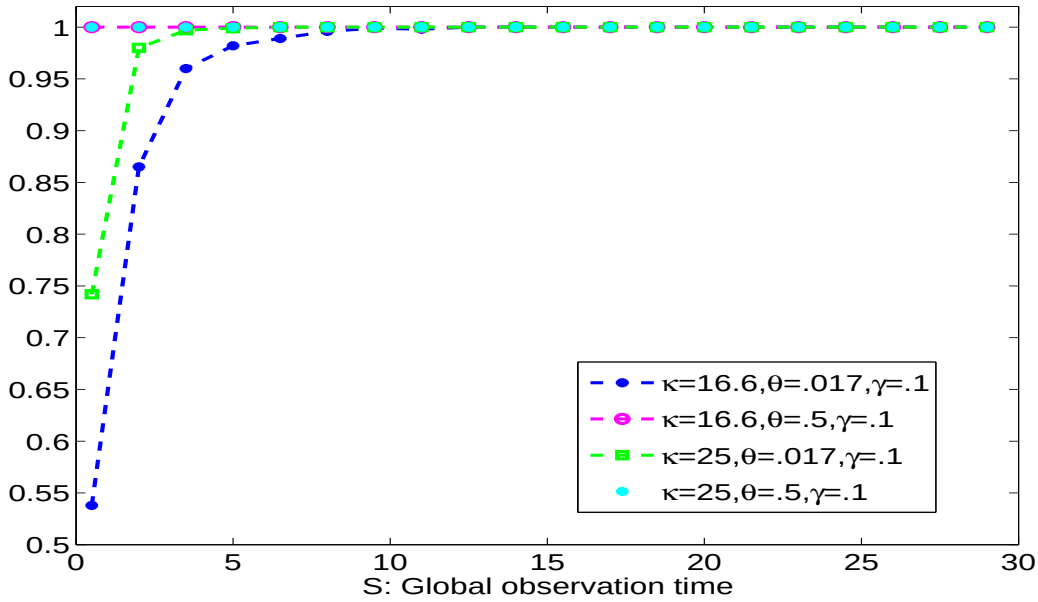


Figure 5.10: Convergence to 1 of Pr_θ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1,000$ and $\gamma = .1$.

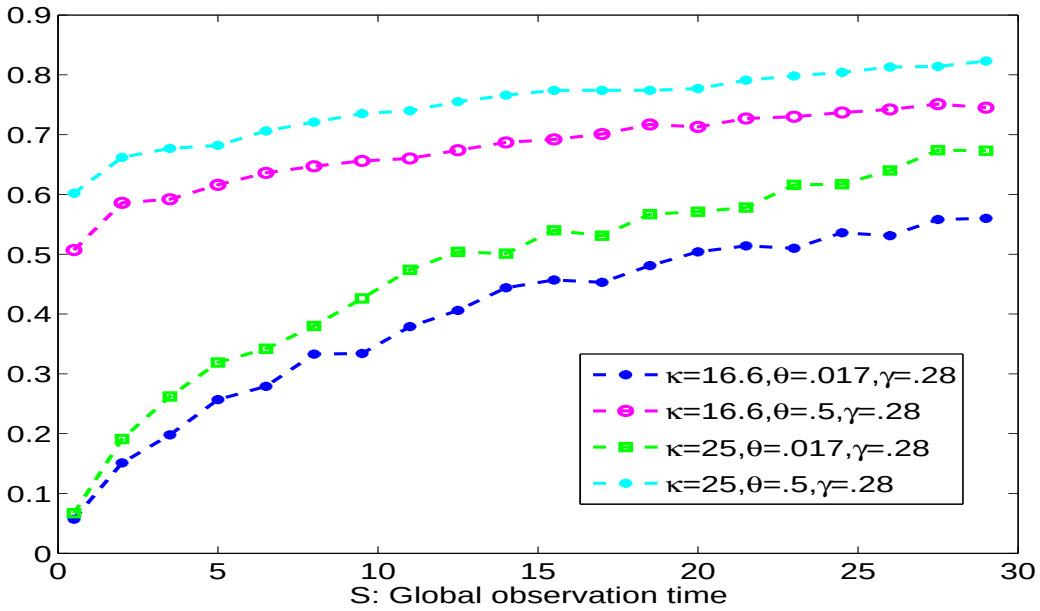


Figure 5.11: Convergence to 1 of Pr_κ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1000$ and $\gamma = .28$.

Fig. 5.11, Fig. 5.12 illustrate similar results for the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ but now for γ fixed at $\gamma = .28$. Again we observe that for our 4 benchmark models, we have $Pr_\gamma = 1$ as soon as $S \geq 10$, which corresponds to $N \geq 10,000$ and so we do not display the results for $\hat{\gamma}$. We note that the Pr_κ and Pr_θ converge to 1 for all these 4 benchmark models. We observe that for higher values of γ we need larger numbers N of observations to reach adequately high values for Pr_κ and Pr_θ.

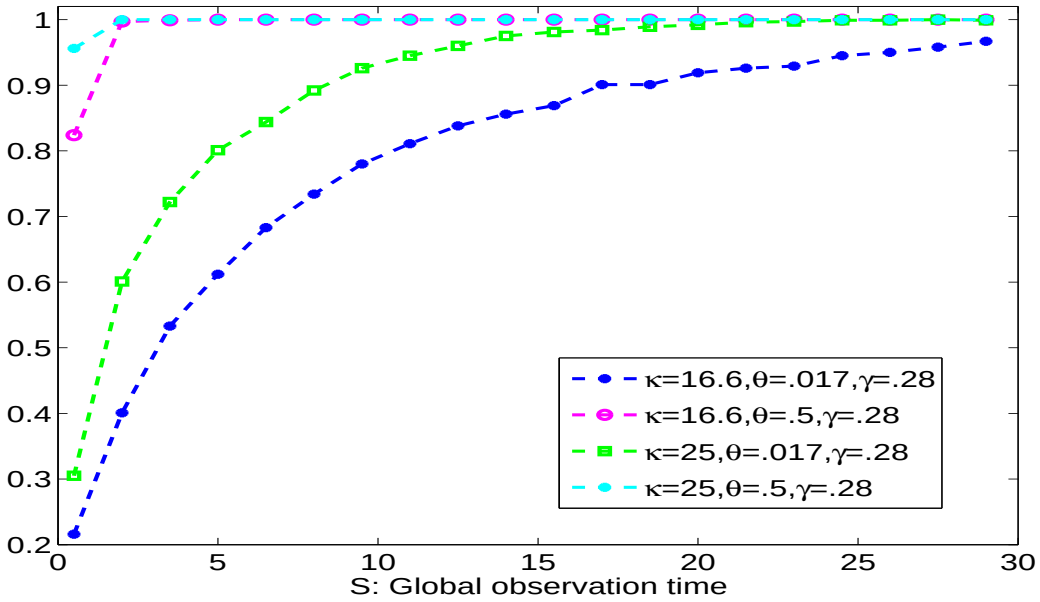


Figure 5.12: Convergence to 1 of Pr_θ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1000$ and $\gamma = .28$.

5.7 Asymptotic behavior of $\hat{\rho}$ and $\hat{\mu}$

We now present numerical results on the convergence of the probabilities,

$$Pr_\rho(S,T) = Pr(|\hat{\rho} - \rho| < tol_\rho), \quad Pr_\mu(S,T) = Pr(|\hat{\mu} - \mu| < tol_\mu),$$

at the tolerance levels given below,

$$tol_{(\rho=-.54)} = .03, \quad tol_{(\mu=.3)} = .045, \quad tol_{(\rho=-.8)} = .04, \quad tol_{(\mu=.6)} = .09.$$

We present our results for the following 4 benchmark Heston models,

$$\begin{array}{cc} \rho & \mu \\ -.54 & .3 \\ -.8 & .6 \\ -.54 & .6 \\ -.8 & .3 \end{array} \tag{5.5}$$

with κ, θ, γ fixed at $\kappa = 16.6, \theta = .017$ and $\gamma = .28$.

Fig. 5.13 and Fig. 5.14 display the convergence to 1 of Pr_ρ and Pr_μ for fixed $T = 1/1000$ when S increases from 0.5 to 30 by increments of 1.5. This corresponds to values of N increasing from 500 to $30,000$ by increments of 1,500.

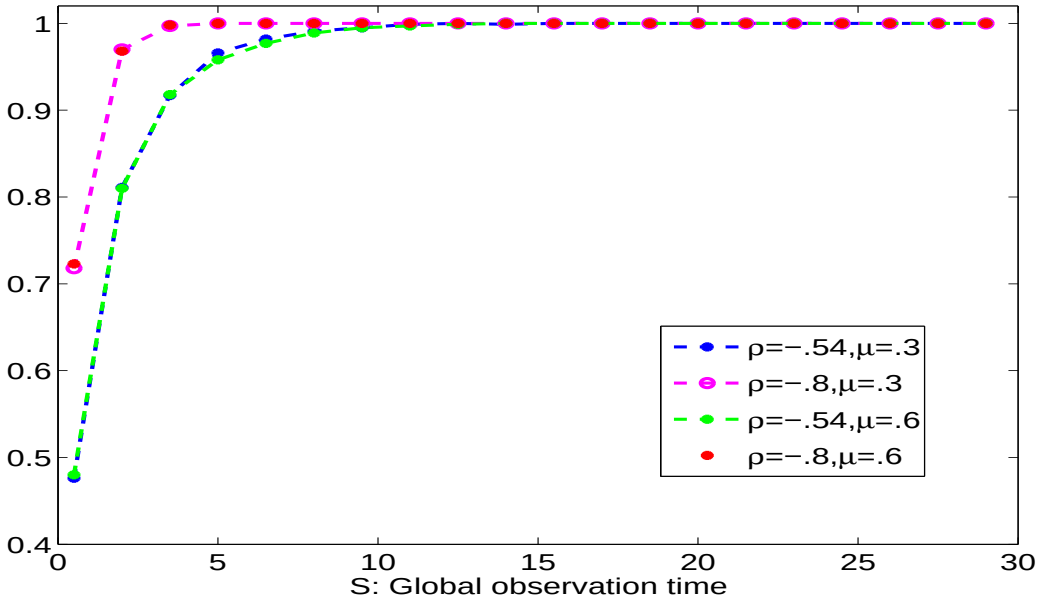


Figure 5.13: Convergence to 1 of Pr_ρ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1000$, $\kappa = 16.6, \theta = .017, \gamma = .28$.

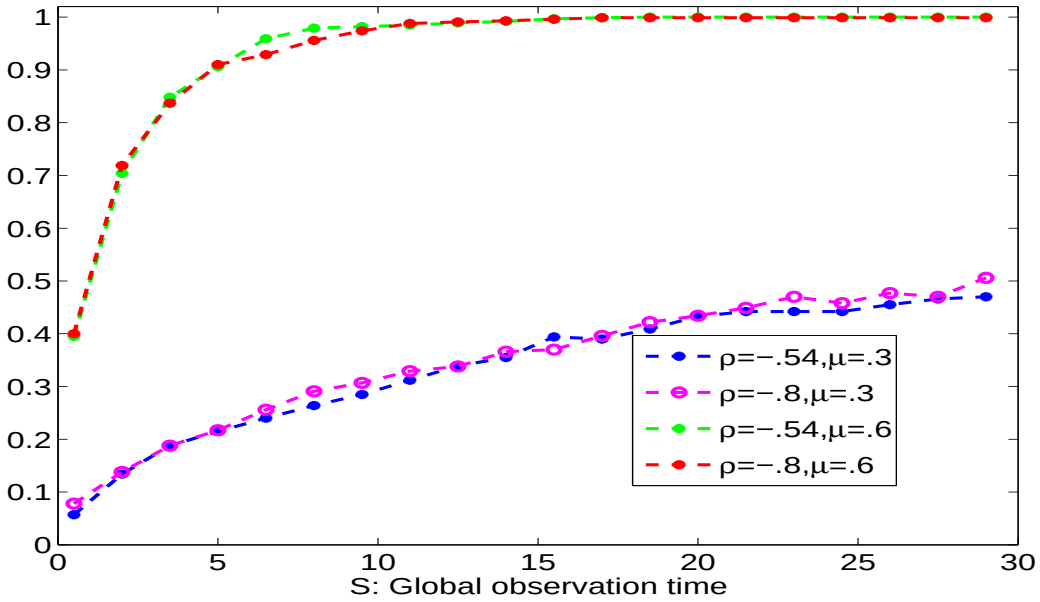


Figure 5.14: Convergence to 1 of Pr_μ as $N = S/T$ increases from 500 to 30,000. Here $T = 1/1000$, $\kappa = 16.6, \theta = .017, \gamma = .28$.

5.8 Variance and covariance of estimators

We show numerically below that the covariance matrix of $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ and $\hat{\rho}$ corresponding to the parameter vector P of the Heston model HSP model (row one of (5.4)) can be approximated by a deterministic symmetric positive definite matrix $L(P)/N$ for a reasonable range of N values and small T fixed. In view of the sensitivity study further below, we will fix $T = 1/252$.

We compute the empirical variances and covariances of our parameter estimators when S increases from 0.5 to 60.5 by increments of 1.5. Our empirical estimates are computed after simulating 1000 random trajectories, using a very small discretization step δ as above. The corresponding number of observations N increases from 500 to $15,246$ by increments of 1500.

Fig. 5.15 shows that for this range of N values the empirical variance of all the four main parameters is very small.
Fig. 5.16 shows that in the same range of N-values, the empirical covariances between these 4 parameter estimators are practically negligible.

Let Σ_N denote the 4×4 empirical covariance matrix of the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}, \hat{\rho}$ when one has N observations. We show in Fig. 5.17 that for N large enough, the matrix $N \times \Sigma_N$ is well approximated by a fixed diagonal matrix. In large classes of situations, this type of asymptotic result is often valid for classical maximum likelihood estimators, but of course the limit matrix is not necessarily diagonal [5]. In our numerical simulations, we estimate Σ_N by computing empirical covariances over 1000 simulated random trajectories.

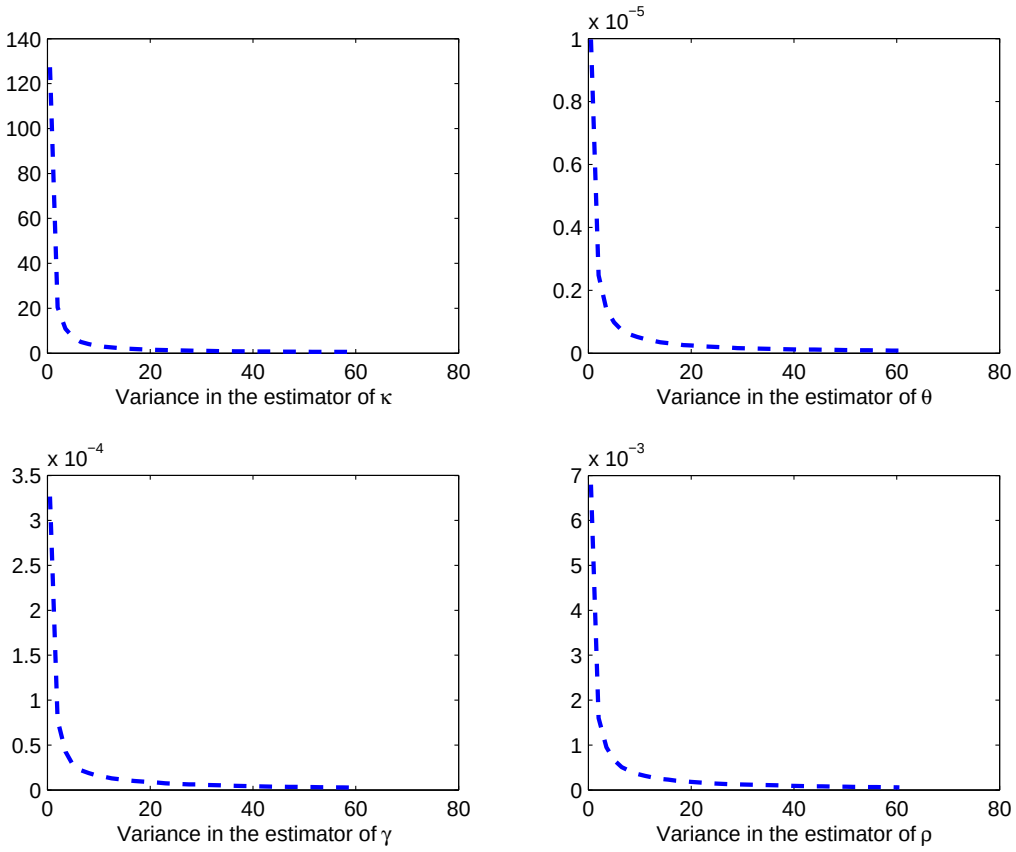


Figure 5.15: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28,\ \rho = -.54,\ \mu = .01$. Decreasing variances of 4 parameter estimators. The abscissa is the global observation time S. The number of observations is $252 \times S$.

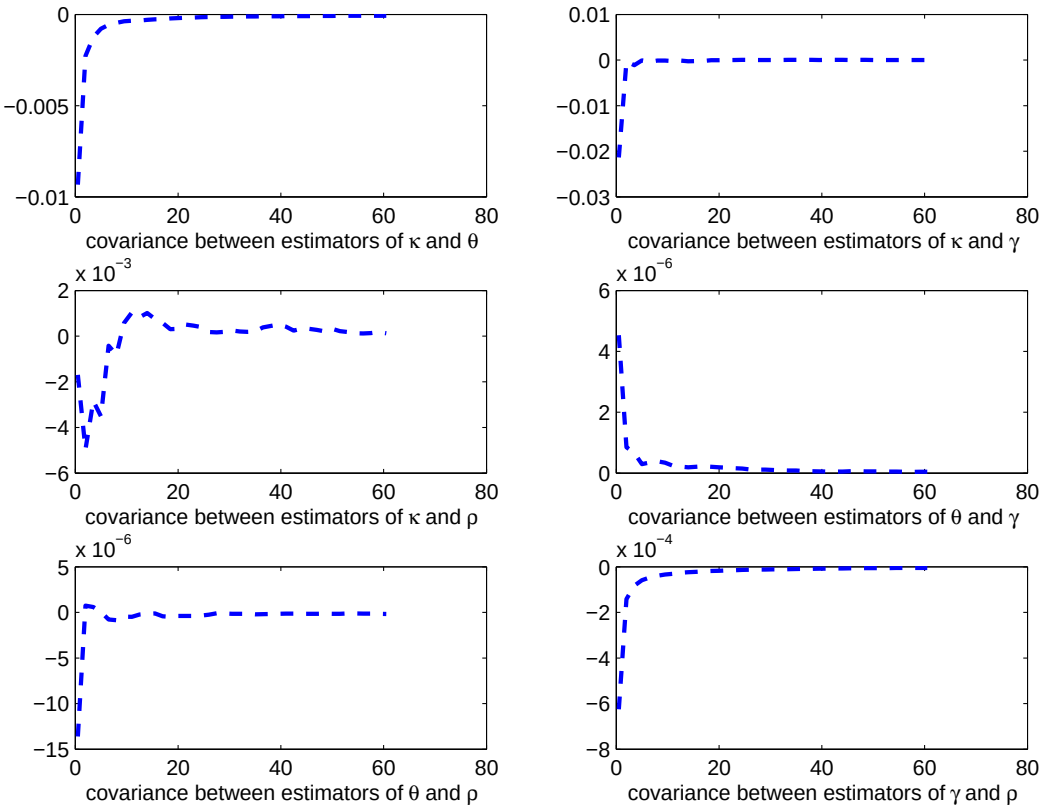


Figure 5.16: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28$, $\rho = -.54$, $\mu = .01$. Negligible covariances of 4 parameter estimators. The abscissa is the global observation time S. The number of observations is $252 \times S$.

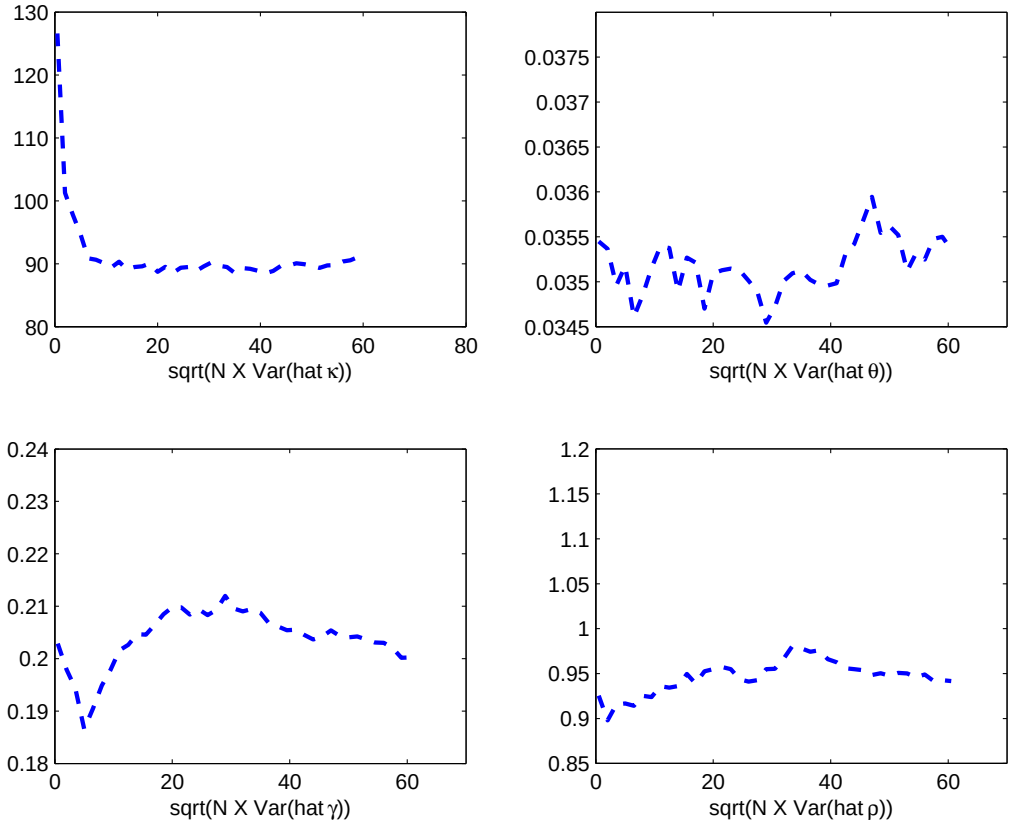


Figure 5.17: Model Parameters : $\kappa = 16.6, \theta = .017, \gamma = .28,\ \rho = -.54,\ \mu = .01.$ Asymptotic stabilization of $N\times$ (variance) for our 4 parameter estimators . The abscissa is the global observation time S. The number of observations is $252 \times S$

Let P denote the vector of parameters for our Heston models, and call $\hat{P}$ our estimator of P. Call Σ_N the covariance matrix of $\hat{P}$ when N is the number of observations. Denote by $L(P)$ the limit of $N\Sigma_N$ as $N \to \infty$, which is a deterministic positive definite matrix.

When P is the vector of parameters for the Heston model HSP, we observe numerically that for fixed small T, the asymptotic bias of the vector of parameter estimators $\hat{P}$ is very small for the reasonable range of N values considered above. We can then approximate the L_2-norm estimation errors for $\hat{P}$ by the diagonal terms of the matrix $\sqrt{L(P)}/\sqrt{N}$.
For the Heston model HSP, we have obtained the following numerical expression of the matrix $L(P)$ (5.6):

$$\begin{pmatrix} 8.12 & -1.02 & .05 & 2.61 \\ -1.02 & .001 & .001 & -.002 \\ .054 & .001 & .041 & -.075 \\ 2.61 & -.002 & -.075 & .9 \end{pmatrix}. \tag{5.6}$$

We will use this approximation of parameter estimation errors in chapter 8 to study their impact on option pricing for the realistic value of $N = 252$.

The asymptotic covariance matrix corresponding to $N = 252$ for the parameter estimators of the Heston model HSP is then approximated by the matrix $\sqrt{L(P)}/\sqrt{N}$ given by,

$$\begin{pmatrix} 5.67 & -0.001 & 0.000 & 0.002 \\ -0.001 & 0.002 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.012 & -0.004 \\ 0.002 & 0.000 & -0.004 & 0.06 \end{pmatrix}. \tag{5.7}$$

We observe that all the off-diagonal elements of this covariance matrix are negligible.

5.9 Asymptotic prevalence of Case 1 for the estimation algorithms

We now construct a synthetic Heston model where the parameters u, v, w are close to the boundary situation $u = w$. We then show numerically, that even in this near boundary situation, and for a fixed small $T = 1/1000$, then as S becomes large the probability of Ω_S tends to 1. Recall that Ω_S is the set of random trajectories for which the Case 1 estimation algorithmics is applicable.

We thus consider the Heston model HBD parametrized by,

$$\kappa = 2, \quad \theta = .2, \quad \gamma = .85, \quad \rho = -.54, \quad \mu = .10.$$

Then $2\kappa\theta = .8$ and $\gamma^2 = .72$, so that the condition $2\kappa\theta - \gamma^2 > 0$ is narrowly satisfied.

We simulate 1000 trajectories of the heston model HBD with simulation step $\delta = 1/20,000$. We fix $T = 1/1000$ and study $Pr(\Omega_S)$ when S increases from 0.5 to 30 by increments of 0.5. These probabilities are estimated by empirical frequencies over our random sample of 1000 trajectories.

Fig. 5.18 displays these numerical results and validates empirically the fast convergence to 1 of $Pr(\Omega_S)$ when S becomes large.In particular for $S \geq 13$, we practically have $Pr(\Omega_S) = 1$.

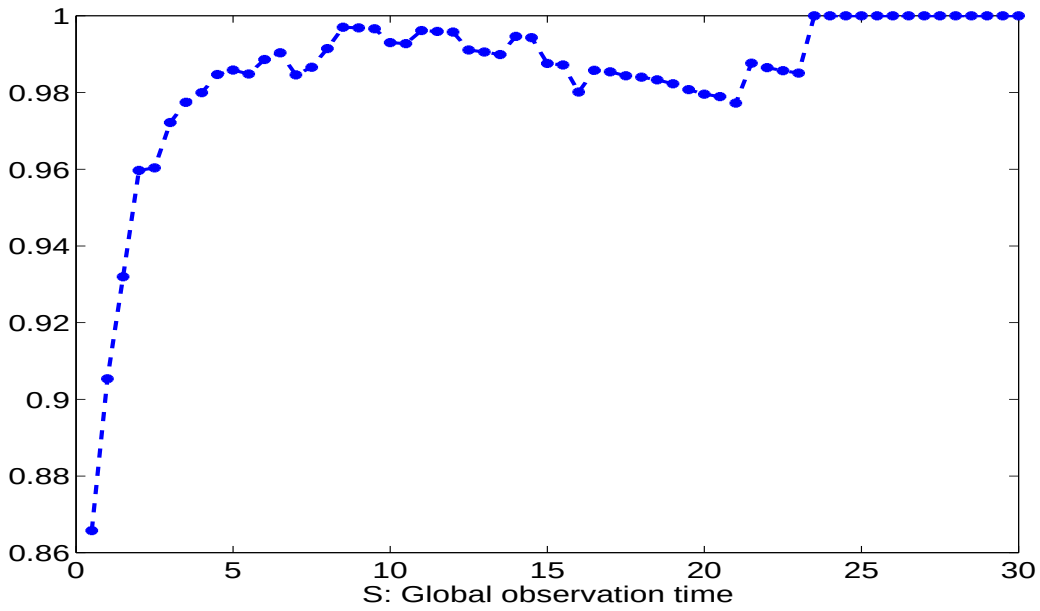


Figure 5.18: As S becomes large $Pr(\Omega_S) \rightarrow 1$

Chapter 6

Option pricing : Introduction and survey of previous approaches

6.1 Introduction

The price of an option when the underlying asset price satisfies a stochastic differential equation can be obtained, under certain assumptions on the market and the functional form of the price (see chapter 7), by solving an associated parabolic partial differential equation (PDE). The underlying model parameters appear as coefficients in the option pricing partial differential equation. The true values of the model parameters are never known. To evaluate option prices through robust model based inference from asset dynamics data, it is therefore crucial that the underlying stochastic dynamics of the asset price be calibrated correctly. We described an efficient approximate maximum likelihood approach to estimate the parameters of Heston's stochastic volatility model in chapter 3. The empirical evaluation of the estimation errors was presented in chapter 5 for a large set of benchmark Heston models. We will now study the impact of estimation errors on the price of options when the underlying asset price satisfies the Heston SDEs. We will give a formal definition

of the option pricing errors in chapter 7. To compute the impact of estimation errors on option price we will obtain and solve the partial differential equations that the derivatives of the option price with respect to the model parameters satisfy.

The option pricing PDE for the case, when the underlying asset price follows a geometric Brownian motion was derived by Black and Scholes [15]. For this case the pricing PDE is derived by constructing a portfolio with the option to be priced and the underlying asset. The idea is to determine a portfolio that is locally deterministic or riskless. Under no arbitrage principles [13], the rate of return on this portfolio has to equal the risk free rate of return. This leads to the option pricing PDE. In the stochastic volatility model, such as Heston's model, there are two sources of randomness driving respectively the asset price and volatility, and only one underlying tradable asset. It is not possible then to eliminate the risk completely by considering a portfolio consisting only of the asset price and the option to be priced. Therefore a second option written on the same underlying asset is added to the portfolio. The no arbitrage principle introduces a new parameter, called the market price of volatility risk, on which the option pricing PDE then depends. The articles [69] and [36] derive the option pricing PDE under the Heston model by constructing a locally riskless portfolio. A first attempt to estimate the market price of volatility risk from observed options data for the S&P 500 options is presented in chapter 6. We study the sensitivity of the option price with respect to this new parameter also and present numerical results on its estimation and sensitivity in chapter 8.

For the price of an European option, Heston [44] gives a semi-closed form in terms of the inverse characteristic function of the underlying probability distributions (section 7.4). Carr and Madan [17] propose a Fast Fourier Transform method to solve for the European option price in this form. Quadrature methods for numerical integration of the inverse characteristic function are discussed in [61]. Alternatively, the option pricing PDE can be

directly solved using numerical methods for partial differential equations (PDEs) [1]. We use a finite difference scheme in space and the backward difference formula in time to solve the PDEs satisfied by the option price and its derivatives with respect to the parameters (chapter 8). The finite difference scheme we use follows closely the discretization method proposed by Ikonen and Toivanen [49]. At the boundary where volatility equals zero we use an upwind scheme [41].

We will illustrate by detailed numerical examples in chapter 8 the option pricing errors for four options with actively traded strike prices in the first quarter of 2007 under four benchmark Heston models. We estimated the Heston model for the 2006 S&P 500 data in chapter 5 using our decoupled estimation method. We now observe the SPX and VIX values for the first quarter of 2007 and present results for this realistic range of the asset price and volatility.

6.2 Option Pricing

6.2.1 Market model

In our context we define the financial market to comprise of a risk-free asset price process $\{R_t\}$ and a risky asset price process $\{X_t\}$. The dynamics of the risk-free asset price is deterministic and given by,

$$dR_t = r(t)R_t dt, \tag{6.1}$$

where $r(t)$ is a deterministic function of time called the risk-free rate of return. In our study $r(t) = r$ will be a constant. Therefore,

$$dR_t = rR_t dt, \quad R_t = e^{rt}R_0. \tag{6.2}$$

Let Y_t be the square of the instantaneous volatility of $\{X_t\}$. We assume that the pair $\{X_t, Y_t\}$ satisfies the stochastic volatility model of Heston. We will continue to use the notations for asset price and volatility from our earlier chapters. We recall below the Heston model SDEs,

$$dX_t = \mu X_t dt + \sqrt{Y_t} X_t dZ_t, \tag{6.3}$$

$$dY_t = \kappa(\theta - Y_t)dt + \gamma\sqrt{Y_t} dB_t, \tag{6.4}$$

where the processes Z and B are standard Brownian motions with $E[dZ_t dB_t] = \rho dt$. Recall, that the natural domain for the parameters is given by,

$$\kappa > 0,\, \theta > 0,\, \gamma > 0, \quad 0 < \theta < 1, \quad -1 \leq \rho \leq 1, \quad 2\kappa\theta > \gamma^2. \tag{6.5}$$

In our numerical computation of the option price we will perform a change of variables to consider the option price as a function of the logarithm of the stock price, denoted by, $LS_t = \log(X_t)$. We state the Ito's lemma [48] below as applicable to our SDEs and then use it to compute the dynamics of LS_t. Let $C^{2,2,1}$ denote the space of functions $h(x, y, t) : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ such that h is continuously differentiable in t and twice continuously differentiable in x and y.

Lemma 6.2.1. Ito's lemma *: For any $C^{2,2,1}$ function h the stochastic differential equation of $h(X_t, Y_t, t)$ where (X_t, Y_t) satisfy* (6.3)-(6.4) *is given by*

$$\begin{aligned} dh(X_t, Y_t, t) = \Bigg[\frac{\partial h}{\partial t}dt + \mu x\frac{\partial h}{\partial x} + \kappa(\theta - y)\frac{\partial h}{\partial y} + \frac{1}{2}x^2 y\frac{\partial^2 h}{\partial x^2} + \frac{1}{2}\gamma^2 y\frac{\partial^2 h}{\partial y^2} + \dots \\ \rho\gamma xy\frac{\partial^2 h}{\partial x \partial y}\Bigg] dt + \mu x\frac{\partial h}{\partial x}dW_1(t) + \kappa(\theta - y)\frac{\partial h}{\partial y}dW_2(t). \end{aligned} \tag{6.6}$$

Proposition 6.2.1. *Let X_t satisfy the SDE* (6.3)*. The dynamics of $LS_t = \log(X_t)$ is given*

by,

$$dLS_t = (\mu - \frac{1}{2}Y_t)dt + \sqrt{Y_t}dB_t. \tag{6.7}$$

The proof follows from a direct application of Ito's lemma (6.6).

6.2.2 Option pricing PDE

Definition 6.2.2. *A European call option with strike (exercise) price K and time to maturity (exercise date) T_M on an underlying asset with price process $\{X_t\}$ is a contract defined by the following clauses,*

- *The holder of the option has at time T_M, the right, but not an obligation, to buy one share of the underlying stock at the price K units from the underwriter of the option.*
- *The right to buy the underlying stock at the price K can only be exercised at the precise time T_M.*

The exercise price K and the time to maturity T_M are determined at the time when the contract is created which we will assume to be $t = 0$. While the contract can be exercised only at maturity, it can be traded at any time from $t = 0$ to $t = T_M$.

The European call option contract with strike price K and maturity T_M written on the underlying asset with price process $\{X_t\}$ is characterized by its pay-off function $\Phi : \mathbb{R}^+ \to \mathbb{R}^+ \cup \{0\}$,

$$\Phi(X_{T_M}) = (X_{T_M} - K)^+ = \max(X_{T_M} - K, 0),$$

where X_{T_M} is the price of the asset at time T_M. We will study the pricing of this option contract under the following standard assumptions on the market:

(1) The market is free of arbitrage, i.e., there is no possibility of making a profit out of nothing without any risk.

(2) The market is frictionless, i.e., there are no transaction costs.

(3) There is unlimited liquidity in the market, i.e., the option can be bought and sold at any time.

Let $\{Z_t, 0 \leq t \leq T_M\}$ be the price of an European call option with maturity T_M written on an underlying asset whose price process satisfies the Heston model (6.3)-(6.4). It is standard to assume that under stochastic volatility the price Z_t is a function of X_t and Y_t [44],

$$Z_t = f(X_t, Y_t, t).$$

Standard arbitrage arguments [15, 58, 44] demonstrate that the value of the option $f(X_t, Y_t, t)$ under sufficient regularity assumptions, must satisfy the following partial differential equation (PDE) on the domain $\mathbb{R}^+ \times \mathbb{R}^+ \times [0, T_M]$,

$$\frac{\partial f}{\partial t}(x,y,t) + rx\frac{\partial f}{\partial x}(x,y,t) + [\kappa(\theta - y) - \lambda_t\sqrt{y}\gamma]\frac{\partial f}{\partial y}(x,y,t) + \ldots$$
$$\frac{1}{2}x^2 y\frac{\partial^2 f}{\partial x^2}(x,y,t) + \frac{1}{2}\gamma^2 y\frac{\partial^2 f}{\partial y^2}(x,y,t) + \rho\gamma xy\frac{\partial^2 f}{\partial x \partial y}(x,y,t) - rf(x,y,t) = 0. \tag{6.8}$$
$$f(x,y,T_M) = \Phi(x). \tag{6.9}$$

where r is the risk free rate of return from (6.2). Note that the pricing PDE does not depend on the parameter μ. The terminal condition $f(x, y, T_M) = \Phi(x)$ is explained by the following simple argument. If at time T_M, $X_T \geq K$, we can make a profit of $(X_{T_M} - K)$ by exercising the option. If $X_{T_M} < K$ the option has no value whatsoever. Therefore, the only reasonable price of the option at time T_M is $\max(X_{T_M} - K, 0) = \Phi(X_{T_M})$.

The unspecified term λ_t that appears in (6.8) is a random variable called the *market price of volatility risk* and arises due to the non tradability of the volatility [13]. We will assume λ_t to be a constant denoted by λ similarly to [33] and [44]. We will sketch in section 6.2.3 a method to estimate λ and study in subsequent chapters the sensitivity of option price to this new parameter.

The option pricing PDE (6.8)-(6.9) is a final value problem. By a transformation of time $t \to T_M - t$ we formulate the problem as an initial value problem. Define

$$g(x, y, t) = f(x, y, T_M - t), \quad 0 \leq t \leq T_M.$$

This immediately implies from (6.8)-(6.9)

$$\frac{\partial g}{\partial t} - \mathcal{L}g = 0 \text{ on } (0, \infty) \times (0, \infty) \times (0, T_M], \tag{6.10}$$

with initial condition,

$$g(x, y, 0) = (x - K)^+ \text{ on } (0, \infty) \times (0, \infty) \times \{t = 0\}, \tag{6.11}$$

where $\mathcal{L}$ is the linear operator,

$$\mathcal{L} = rx\frac{\partial}{\partial x} + [\kappa(\theta - y) - \lambda\sqrt{y}\gamma]\frac{\partial}{\partial y} + \frac{1}{2}x^2 y\frac{\partial^2}{\partial x^2} + \ldots$$
$$\frac{1}{2}\gamma^2 y\frac{\partial^2}{\partial y^2} + \rho\gamma xy\frac{\partial^2}{\partial x \partial y} - r.$$

Note that we are now assuming a constant market price of volatility risk. We now describe the boundary conditions satisfied by the option price [44]. The boundary conditions remain unchanged after the transformation in the time variable. When the asset price is zero there

is no rational interest in buying the call option and therefore the option is worthless,

$$g(0, y, t) = 0 \text{ on } \{x = 0\} \times (0, \infty) \times (0, T_M]. \tag{6.12}$$

For large values of the asset the option price grows linearly with the asset price,

$$\lim_{x \to \infty} \frac{\partial g(x, y, t)}{\partial x} = 1, \forall y,\ t. \tag{6.13}$$

For large values of Y_t the option price tends to be constant as a function of the square of volatility,

$$\lim_{y \to \infty} \frac{\partial g(x, y, t)}{\partial y} = 0, \forall x,\ t. \tag{6.14}$$

The boundary condition at $y = 0$ is obtained by setting y to be zero in the PDE (6.10),

$$\frac{\partial g}{\partial t} - rx\frac{\partial g}{\partial x} - \kappa\theta\frac{\partial g}{\partial y} + rg = 0, \tag{6.15}$$

on $(0, \infty) \times \{y = 0\} \times (0, T_M]$. For all practical applications we can consider the problem on the bounded domain,

$$U_{T_M} = U_B \times (0, T_M],$$

where

$$U_B = (0, X_{max}) \times (0, Y_{max}).$$

The option pricing problem is a two-dimensional second-order parabolic partial differential equation where the operator L is elliptic for each t in x and y [34]. More precisely it is a convection-diffusion type equation with Dirichlet boundary at $x = 0$ and Neumann

boundary as x and y approach X_{max} and Y_{max} respectively. The boundary condition at $y = 0$ is not of a standard type [40]. We refer to Equation(6.10) together with the initial and boundary conditions (6.11)-(6.15) as an initial/boundary-value problem [30]. The classical conditions under which the initial/boundary value problem has a unique solution are given in [34]. These conditions are not satisfied by the PDE (6.10) on the unbounded spatial domain, in particular uniform ellipticity fails.

Recall that we had imposed the condition $2\kappa\theta > \gamma^2$ on the model parameters. This imposes that the boundary $y = 0$ in not hit. The paper [43] gives sufficient conditions on the coefficients of the underlying SDEs under which the *probabilistic solution* of the option price is the unique classical solution of the option pricing PDE with an initial condition of type (6.11). For the Heston model they show that under the condition $2\kappa\theta > \gamma^2$ (so that the process Y_t stays away from zero with probability one) their sufficiency conditions are satisfied and there is a unique classical solution of the initial/boundary (6.10)-(6.15) value problem given after a transformation of time by

$$f(x, y, t) = e^{-r(T_M - t)} E_{t,x,y}[\Phi(X_{T_M})], \tag{6.16}$$

where $E_{t,x,y}[\ .\ ] = E[\ .\ |X_t = x, Y_t = y]$ and the expectation is with respect to the following dynamics of (X_t, Y_t)

$$dX_t = rX_t dt + \sqrt{Y_t} X_t dW_1(t), \quad X_t = x, \tag{6.17}$$

$$dY_t = \kappa((\theta - Y_t) - \lambda\gamma\sqrt{Y_t})dt + \gamma\sqrt{Y_t} dW_2(t), \quad Y_t = y, \tag{6.18}$$

where W_1, W_2 are standard Brownian motions with $E[dW_1(t)dW_2(t)] = \rho dt$. The formula (6.16) is called the *risk neutral valuation* of the option price (see [42] for the theory of risk neutral option pricing). Since the joint density of (X_t, Y_t) under the new dynamics

(6.17)-(6.18) is not known we do not have a closed form solution of the option price from its probabilistic representation. Therefore we will compute the option price by solving the initial/boundary value problem numerically.

We note that the SDE satisfied by $LS_t = \log(X_t)$ when X_t satisfies (6.17) is obtained by an application of Ito's lemma (6.6),

$$dLS_t = (r - \frac{1}{2}Y_t)dt + \sqrt{Y_t}dW_1(t). \tag{6.19}$$

6.2.3 Estimation of the market price of volatility risk λ

We need an estimate of λ in order to solve the pricing PDE (6.10). In order to avoid arbitrage possibilities in the market, the value of λ is determined from one benchmark asset and used to price all other assets [13]. The value of λ depends in a non-trivial way on various economic factors prevalent in the market such as liquidity concerns, risk aversion and the utility preferences of investors [13]. Therefore, the problem of determining λ is not a theoretical one but an empirical one. For equity options [54] presents evidence of non-zero market price of volatility risk. Among the existing literature on this subject, [33] uses and estimates a constant λ computed from observed option prices.

We give below a methodology to estimate the market price of volatility risk from options data. Our objective is to get a reasonable estimate of λ such that the theoretical option price computed from the pricing PDE (6.10) is close to the true option price. We use a Black-Scholes approximation of the option price to compute λ.

Definition 6.2.3. *The value of an option, $F = F(S_t, t)$ in the Black-Scholes model satisfies the PDE*

$$\frac{\partial F}{\partial t} + rs\frac{\partial F}{\partial x} + \frac{1}{2}s^2\sigma^2\frac{\partial^2 F}{\partial s^2} - rF = 0, \tag{6.20}$$

where the dynamics of the underlying asset price S_t is given by,

$$dS_t = \mu_F S_t dt + \sigma S_t dW_t. \tag{6.21}$$

Definition 6.2.4. *Let σ_{BS} be the Black-Scholes Implied Volatility corresponding to the observed price OPT of an option on the asset S at time t (with expiration T_1 and strike price K_1) in the market. Then by definition for $\sigma = \sigma_{BS}$ the theoretical solution of the Black-Scholes PDE* (6.20) *is equal to OPT. Note that the implied volatility σ_{BS} is unique by the inverse function theorem because the Black-Scholes option price is a strictly increasing function of volatility [15].*

We describe below as an algorithm the estimation of λ at any time $t < T_M$.

(i) Let $F = F(X_t, t, \sigma)$ be the observed price of an option at a fixed time t in the market with strike K and maturity time T_M. Let σ_{BS} denote the Black-Scholes implied volatility of this option at time t. Then by definition F satisfies:

$$\frac{\partial F}{\partial t} + rx\frac{\partial F}{\partial x} + \frac{1}{2}x^2\sigma_{BS}^2\frac{\partial^2 F}{\partial x^2} - rF = 0, \tag{6.22}$$

together with the boundary condition $F(X_{T_M}, T_M, \sigma) = (X_{T_M} - K)^+$. The solution for the PDE (6.22) is given by [13],

$$F(t, x, \sigma_{BS}) = xN(d_1(t, x, \sigma_{BS})) - e^{-r(T_M - t)}KN(d_2(t, x, \sigma_{BS})). \tag{6.23}$$

where N is the cumulative distribution of a standard Gaussian random variable and

$$d_1(t, x, \sigma_{BS}) = \frac{1}{\sigma_{BS}\sqrt{T_M - t}}\left\{\log(\frac{x}{K}) + (r + \frac{1}{2}\sigma_{BS}^2)(T_M - t)\right\},$$

$$d_2(t, x, \sigma_{BS}) = d_1(t, x, \sigma_{BS}) - \sigma_{BS}\sqrt{T_M - t}.$$

(ii) We want to estimate λ_t so that the solution of the option pricing PDE (6.8)-(6.9) is close to F_t. Using (6.22), this implies that we need to determine λ_t so that F satisfies the following PDE,

$$[\kappa(\theta - y) - \lambda_t\sqrt{y}\gamma]\frac{\partial F}{\partial y} + \frac{1}{2}x^2(y - \sigma_{BS}^2)\frac{\partial^2 F}{\partial x^2} + \frac{1}{2}\gamma^2 y\frac{\partial^2 F}{\partial y^2} + \rho\gamma xy\frac{\partial^2 F}{\partial x \partial y} = 0, \quad (6.24)$$

We assume that in a neighborhood of σ_{BS}, the price of the option F is well approximated by the Black-Scholes formula, i.e.,

$$F(x, y, t) = xN(d_1(x, y, t)) - e^{-r(T_M - t)}KN(d_2(x, y, t)). \quad (6.25)$$

for y in a small neighborhood of σ_{BS}.

(iii) We compute now the partial derivatives of F in this neighborhood and then set $\sigma_{BS} = y$. We solve the resulting equation to get

$$\lambda_t = \frac{\left[\kappa(\theta - \sigma_{BS}^2) + \frac{\gamma^2}{4}(d_1^2 - d_1\sigma_{BS}\sqrt{(T_M - t)} - 1) + \rho\gamma\sigma_{BS}^2(1 - d_1/\sigma_{BS}\sqrt{(T_M - t)})\right]}{\gamma\sigma_{BS}}. \quad (6.26)$$

Therefore, our choice of λ_t is such that, given an option, and its Black-Scholes implied volatility, σ_{BS}, the solution of the option price under the SV model is equal to the price of the option under the Black-Scholes model at $y = \sigma_{BS}$. We have applied this method to estimate λ using options written on the S&P 500 index. We report results for this method and compare it to a best fit method in chapter 8.

Chapter 7

Option price sensitivity to errors in stochastic modeling

7.1 Impact of model estimation errors on option price

We compute a bound on the error in the option price generated by random errors on the parameters κ, θ, γ and ρ of the joint Heston SDEs. These random errors are due to the estimation of the model parameters from observed data on asset price and volatility. In chapter 3 we described the estimation of parameters using a constrained approximate maximum likelihood approach. The estimation error for a parameter will be defined as the L_2 norm of the difference between the estimator and the true parameter value.

The option price $g(x, y, t)$ as a solution of the PDE (6.10), depends on all the model parameters except μ. We will use both an extended and a shorthand notation for g, defined by

$$g(x, y, t; H, \lambda) = g(H) = g(x, y, t),$$

where $H = (H_i)_{i=1}^4$ is the vector,

$$H_1 = \kappa, \quad H_2 = \theta, \quad H_3 = \gamma, \quad H_4 = \rho.$$

It can be proved that $g(H)$ is a smooth function of the parameter vector H in the domain $\mathcal{C}_H$ defined above by Equation (6.5). We will sketch a proof for this in section 7.4.

7.2 Confidence neighborhood of estimators

Let P be the true but unknown value of the parameter vector H. We expect P to be in a neighborhood of the parameter estimator $Q = \hat{P}_N$, where N is a fixed number of daily observations of the pair (asset price, volatility) used to compute $Q = \hat{P}_N$. Recall that the estimators $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ and $\hat{\rho}$ also depend on the value of the time T between consecutive observations. For the sensitivity study presented, T will be a fixed suitably small value.

For all $i = 1, 2, 3, 4$ define the partial derivatives of the option price $g(P)$ at P by

$$D_{i,P} = D_i(g)(x, y, t; H, \lambda) = \frac{\partial g}{\partial H_i}(x, y, t; H, \lambda)$$

and denote by D_P the corresponding gradient of $g(P)$, viewed as a column vector in $\mathbb{R}^4$,

$$D_P = \begin{pmatrix} D_1(g) \\ D_2(g) \\ D_3(g) \\ D_4(g) \end{pmatrix} = \begin{pmatrix} \frac{\partial g}{\partial H_1} \\ \frac{\partial g}{\partial H_2} \\ \frac{\partial g}{\partial H_3} \\ \frac{\partial g}{\partial H_4} \end{pmatrix},$$

where we have suppressed the dependence of the partial derivatives on $(x, y, t; H, \lambda)$. A first-order Taylor expansion of $g(H)$ at $H = P$ under the assumption that the second-order

partial derivatives of $g(H)$ with respect to the parameters κ, θ, γ and ρ are continuous yields

$$g(Q) \simeq g(P) + \sum_{i=1}^{4} D_{i,P}.(Q_i - P_i). \tag{7.1}$$

The covariance matrix of the estimator Q denoted by $Cov_{Q,P} = \Sigma_N = (\sigma_{i,j})_{i,j=1}^4$ is defined as follows

$$\sigma_{i,i}^2 = E[(Q_i - P_i)^2],\ i = 1, 2, 3, 4, \tag{7.2}$$

$$\sigma_{i,j} = E[(Q_i - P_i)(Q_j - P_j)],\ i, j = 1, 2, 3, 4;\ i \neq j. \tag{7.3}$$

Note that each element of the matrix $Cov_{Q,P}$ depends on N, but we suppress this dependence for brevity. We verified numerically in chapter 5, that (for fixed small T), the covariance matrix $NCov_{Q,P}$ can be approximated for a reasonable range of N values by $L(P)$, where $L(P)$ is a deterministic symmetric non negative matrix that does not depend on N but only on the true parameter values P. We estimated empirically by intensive simulations the matrix $L(P)$ for the parameter vector P corresponding to the 2006 S&P 500 model HSP (5.6).

The option price error induced by errors in parameter estimation has a mean quadratic error which depends on P and is defined by

$$\delta(P)^2 = E[(g(Q) - g(P))^2]. \tag{7.4}$$

Squaring both sides of (7.1) and taking expected values we get,

$$E[(g(Q) - g(P))^2] \simeq \sum_{i=1}^{4} D_{i,P}^2.E[(Q_i - P_i)^2] + \sum_{i \neq j} D_{i,P} D_{j,P}.E[(Q_i - P_i)(Q_j - P_j)]. \tag{7.5}$$

The expected values in the preceding equations are taken with respect to the underlying probability measure Pr. We observed in chapter 5, that for our estimators the empirical covariance terms were negligible for the N values considered there. The numerical results in chapter 5 showed that even for small values of N the covariance between that parameters $\hat{\kappa}, \hat{\theta}, \hat{\gamma}$ and $\hat{\rho}$ were negligible. Therefore the option price error due to estimation errors is simply the sum of the squared L_2 errors of each of the four parameters. For large N, the option price error is approximated by the following formula,

$$\delta(P)^2 \simeq \frac{1}{N} D_P^* L(P) D_P. \tag{7.6}$$

The true P is unknown, but given the estimated Q, we expect that with high probability, the unknown P will belong to a "confidence neighborhood" of Q, denoted $B(Q) \subset \mathcal{C}_H \subset \mathbb{R}^4$. This was illustrated in details, for large N, and very small values of T, for eight different Heston models in chapter (5). For instance, we observed from the numerical examples that the estimators, $\hat{\theta}, \hat{\gamma}$ and $\hat{\rho}$ of the 2006 S&P model were within 5% of the true parameter value with probability greater that 90% at $T = 1/1000$ and $N \geq 20000$.

When P varies within $B(Q)$, the correct option price $g(P)$ and the option price $g(Q)$ computed from the estimated Heston model differ by a random error of size $|g(Q) - g(P)|$, which has a deterministic L_2 error size $\delta(P)$ computed by (7.6). To evaluate the sensitivity of the computed option price to errors in the estimated parameter values Q, we need to compute the maximum of $\delta(P)$ which we denote by $\varepsilon(Q)$ for arbitrary $P \in B(Q)$. A typical "best" choice of $B(Q)$ would be the ellipsoid centered at Q and defined by the covariance matrix Σ_N. For large N, Σ_N is equivalent to $(1/N)L(P)$, and the off diagonal entries of the covariance matrix $(1/N)L(P)$ are negligible. This was show for the model HSP in chapter

5. Hence, for practical applications, we define the confidence neighborhood $B(Q)$ by

$$B(Q) = B_1 \times B_2 \times B_3 \times B_4,$$

with

$$B_i = [Q_i - \frac{1}{\sqrt{N}}\sqrt{L(P)_{i,i}}, Q_i + \frac{1}{\sqrt{N}}\sqrt{L(P)_{i,i}}],$$

where $\frac{1}{\sqrt{N}}\sqrt{L(P)_{i,i}}$ is an estimate of the standard deviation $\sigma_{i,i}$ of Q_i for $i = 1, 2, 3, 4$. To improve the level of confidence for the neighborhood $B(Q)$ one may clearly double the length of the intervals B_i.

7.3 Sensitivity of option price to parametric estimation errors

We then define the *sensitivity* of the option price $g(H)$ with respect to each parameter H_i at the estimated value Q_i naturally as the product $|D_{i,P}|\sqrt{E[(Q_i - P_i)^2]}$ of the absolute value of the partial derivative $D_{i,P}$ of $g(H)$ at P with the L_2 error of H_i. With this more explicit notation, the following 4 sensitivities for the option price can be expressed as

$$\begin{aligned} Sen_\kappa &= |D_{1,P}|\sigma_{1,1}, \quad Sen_\theta = |D_{2,P}|\sigma_{2,2}, \\ Sen_\gamma &= |D_{3,P}|\sigma_{3,3}, \quad Sen_\rho = |D_{4,P}|\sigma_{4,4}, \end{aligned} \tag{7.7}$$

where $\sigma_{i,i} \approx \frac{1}{\sqrt{N}}\sqrt{L(P)_{i,i}}$, $i = 1, 2, 3, 4$. We then proceed to compute the partial derivatives of the option price with respect to the model parameters.

7.4 Differentiability of the option price with respect to model parameters

Before computing sensitivities of the option price to parameter estimation errors, we review the question of existence, uniqueness, and differentiability (with respect to parameters) for the solution of the parabolic PDE and boundary value problem defining the option price.

The extensive literature on parabolic PDEs does not seem to cover explicitly the specific initial/boundary value problem that we have here; see for instance the semi-group method [76] or the weak sense solutions approach [68]. So we sketch here a proof for the differentiability of the option price with respect to its parameters.

The option price is a function of the logarithm lsp of the stock price, denoted by

$$h(lsp, y, t) = f(exp(lsp), y, t),$$

where $t \in (0, T_M),\ lsp \in (-\infty, \infty), y \in (0, \infty)$.

Let $Z_t = (LS_t, Y_t)$. We will also use the shorthand notations $z = (lsp, y)$ and $h(lsp, y, t) = h(z, t)$. The equivalent probabilistic expression for the option price at time t from (6.16) can be written in terms of the logarithm of the stock price,

$$h(z, t) = e^{-r(T_M - t)} E(\Phi(Z_{T_M}) | Z_t = z), \tag{7.8}$$

where $z = (lsp, y)$. Recall the definition of the payoff function in terms of lsp,

$$\Phi(z) = \Phi(lsp, y) = \max(e^{lsp} - K, 0).$$

The conditional expectation of the random variable $\Phi(Z_{T_M})$ given $Z_t = z$ in (7.8) is computed with respect to the dynamics of asset price and volatility given in (6.17)-(6.18). Let

$$u(t) = e^{-r(T_M - t)}.$$

Then,

$$h(z,t) = u(t)E(\Phi(Z_{T_M})|Z_t = z) \text{ and}$$

$$h(z,t) = u(t)\int_\Omega q(t,z;T_M,LS)\Phi(Z)dZ, \tag{7.9}$$

where $q(t,z;T_M,LS)$ is the conditional density of the random variable LS_{T_M} given $Z_t = z$. Define the set $\mathcal{D}$ to be the spatial domain,

$$\mathcal{D} = \mathbb{R} \times \mathbb{R}^+.$$

The transition density function $q(t,z;T_M,LS)$ is known to satisfy the Kolmogorov forward equation as a function of T_M and LS [39],

$$(\frac{\partial}{\partial T} - \Delta)q = 0 \text{ on } (t,\infty) \times \mathcal{D}, \tag{7.10}$$

with

$$\lim_{T \to t} q(t,z;T,LS) = \delta_z,$$

and

$$\Delta = - < b, \nabla > + \frac{1}{2} tr(\mathcal{H}\sigma\sigma^*),$$

where b is the column vector of drift coefficients and σ is the column vector of diffusion coefficients of (6.19) and (6.18). The gradient vector is denoted by ∇ and the trace of the matrix $\mathcal{H}\sigma\sigma^*$ is given by $tr(\mathcal{H}\sigma\sigma^*)$ where $\mathcal{H}$ is the Hessian matrix. In Equation(7.10) above we have omitted the dependence of q on the state and time variables.

It is well known that the solution q of the forward equation is a smooth function of the parameters in (7.10)([34],[46]). To compute the derivative of the option price with respect to the model parameters we need to differentiate with respect to the parameters under the integral sign on the right hand side of equation (7.9). Since the function Φ does not depend explicitly on the parameters, one only needs to verify that the derivatives of q with respect to the parameters are integrable in the domain $\mathcal{D}$. This requires deriving good upper bounds on the derivatives at (lsp, y, t) when y tends to zero, which can be done by applications of the maximum principle (see [30]).

7.5 Analytical solution of option pricing for Heston model

In the context of differentiability we give below the expression of the option price h in a semi-closed form from Heston's paper [44],

$$h(lsp, y, t) = e^{lsp} P_1 + K e^{-r(T_M - t)} P_2,$$

where P_1 and P_2 are probabilities that are expressed as the inverse of their characteristic functions,

$$P_j(lsp, y, T_M; \log(K)) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{e^{-i\varphi \log(K)} c_j(lsp, y, T_M; \varphi)}{i\varphi} \right] d\varphi. \tag{7.11}$$

The characteristic functions c_j are of the following form

$$c_j(lsp, y, t; \varphi) = e^{C_j(T_M - t;\varphi) + D_j(T_M - t, \varphi)y + i\ lsp\varphi},$$

where

$$C_j(\tau, \varphi) = r\varphi i\tau + \frac{a}{\gamma^2}\left\{(b_j - \rho\gamma\varphi i + d_j)\tau - 2\log\left[\frac{1 - g_j e^{d_j\tau}}{1 - g_j}\right]\right\},$$

$$D_j(\tau, \varphi) = \frac{b_j - \rho\gamma\varphi i + d_j}{\gamma^2}\left[\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}}\right],$$

$$g_j = \frac{b_j - \rho\gamma\varphi i + d_j}{b_j - \rho\gamma\varphi i - d_j},$$

$$d_j = \sqrt{(\rho\gamma\varphi i - b_j)^2 - \gamma^2(2u_j\varphi i - \varphi^2)},$$

$$b_1 = \kappa + \lambda - \rho\gamma, \quad b_2 = \kappa + \lambda, \quad a = \kappa\theta, \quad u_1 = .5, \quad u_2 = -.5.$$

The derivatives of the characteristic functions c_j with respect to the parameters κ, θ, γ and ρ exists on the natural domain of the parameters $\mathcal{C}_H$. It can be shown that the resulting integrand in (7.11) decays fast enough for the integrals corresponding to the partial derivatives to exist [51].

7.6 Sensitivity equations

We now include in our study the sensitivity of the option price with respect to λ also. Define the vector $\mathbf{p} \in \mathbb{R}^5$ by $\mathbf{p}_i = H_i$ for $i = 1, 2, \ldots, 4$ and $\mathbf{p}_5 = \lambda$. For all $i = 1, 2, \ldots, 5$ define,

$$D_i(g)(x, y, t; \mathbf{p}) = \frac{\partial g}{\partial p_i}(x, y, t; \mathbf{p}).$$

Differentiating equation (6.10) with respect to each of the parameters p_i the *sensitivity PDEs* take the following form,

$$(\frac{\partial}{\partial t} - \mathcal{L})D_i = G_i(x, y, t; \mathbf{p}) \text{ on } U_T, \tag{7.12}$$

with initial condition

$$D_i(x, y, 0; \mathbf{p}) = 0 \text{ on } U_B \times \{t = 0\}. \tag{7.13}$$

We see that the *sensitivity equations* are two-dimensional non-homogeneous parabolic partial differential equations where the right-hand side is given by,

$$G_1(x, y, t; \mathbf{p}) = (\theta - y)\frac{\partial g}{\partial y},$$

$$G_2(x, y, t; \mathbf{p}) = \kappa\frac{\partial g}{\partial y},$$

$$G_3(x, y, t; \mathbf{p}) = \gamma y\frac{\partial^2 g}{\partial y^2} + \rho xy\frac{\partial^2 g}{\partial x \partial y} - \lambda\sqrt{y}\frac{\partial g}{\partial y},$$

$$G_4(t, x, y; \mathbf{p}) = \gamma xy\frac{\partial^2 g}{\partial x \partial y},$$

$$G_5(t, x, y; \mathbf{p}) = -\gamma\sqrt{y}\frac{\partial g}{\partial y}.$$

The boundary conditions for the sensitivity PDEs are,

$$D_i = 0, \text{ on } \{x = 0\} \times (0, Y_{max}) \times (0, T_M], \text{ with}$$

$$\lim_{x \to X_{max}} \frac{\partial D_i}{\partial x} = 0, \forall y, \ t, \text{ and}$$

$$\lim_{y \to Y_{max}} \frac{\partial D_i}{\partial y} = 0, \forall x, \ t$$

for all $i = 1, 2, \ldots, 5$. For the boundary $y = 0$ we differentiate Equation(6.15) with respect to each of the parameters to obtain,

$$(\frac{\partial}{\partial t} - rx\frac{\partial}{\partial x} - \kappa\theta\frac{\partial}{\partial y} + r)D_i = F_i(x, y, t; \mathbf{p}), \tag{7.14}$$

where

$$F_1 = \theta\frac{\partial g}{\partial y}, \quad F_2 = \kappa\frac{\partial g}{\partial y}$$

and $F_i = 0$ for $i = 3, 4, 5$. We do not have a sensitivity equation corresponding to the data r because r is a known deterministic constant for our purpose and which we do not estimate from the asset price data.

In the next chapter we present the numerical method that is employed to solve the 6 PDEs for the option price and its derivatives with respect to the five parameters.We then present detailed numerical results to evaluate option pricing errors and sensitivity to parameters.

Chapter 8

Option price sensitivity : Numerical study

8.1 Numerical implementation

We solve the option pricing partial differential equations numerically by discretizing the operator $\mathcal{L}$ in (6.12). We use standard schemes for discretization of the option pricing equations for which stability of option price solutions has been shown. We apply a uniform space-time finite difference grid on the computational domain U_{T_M} where

$$U_{T_M} = U_B \times (0, T_M],$$

and

$$U_B = (0, X_{max}) \times (0, Y_{max}).$$

Let the number of grid steps be m, n, and s in the x, y, and t directions respectively. The grid steps in each direction are denoted

$$\Delta x = \frac{X_{max}}{m}; \quad \Delta y = \frac{Y_{max}}{n}; \quad \Delta t = \frac{T_M}{s}.$$

We use superscripts to denote the time variable and subscripts to denote the spatial variable at the grid point values,

$$g_{ij}^k = g(x_i, y_j, t_k) = g(i\Delta x, j\Delta y, k\Delta t),$$

where $i = 0, 1, 2, \ldots, m$; $j = 0, 1, 2, \ldots, n$ and $k = 0, 1, 2, \ldots, s$.

8.1.1 Space discretization

All partial derivatives in the Heston PDE have variable coefficients. In some parts of the domain the first-order spatial derivative terms dominate the second-order terms. The discretization of these spatial derivatives has been considered in [62, 49]. Let $\mathbf{A}$ denote the resulting discretization matrix. If the matrix $\mathbf{A}$ is strictly diagonally dominant with positive diagonal elements and non-positive off diagonal elements (M-matrix) it is known to have good stability properties [71]. We apply the space discretization scheme used in [49] for American options to our European option pricing. We observe that for American options the pricing equation (6.10) is replaced by an inequality but the differential operator $\mathcal{L}$ remains unchanged. The properties of the resulting discretization matrix $\mathbf{A}$ are studied in [49]. In general the matrix $\mathbf{A}$ is not an M-matrix but as remarked in [49], with sufficiently small time steps, the resulting matrix is diagonally dominant. The paper [49] proposes a modified seven point discretization scheme where they artificially increase the size of the second order derivative terms so as to obtain an M-matrix. The discretization scheme with added terms is no longer second-order accurate.

We use a second-order accurate finite difference scheme for the space derivatives. We use the classical central difference scheme for the first-order derivatives and the usual three point scheme for the second-order derivatives. The finite difference operators for the first-order derivatives are,

$$\delta_x g_{i,j}^k = \frac{g_{i+1,j}^k - g_{i-1,j}^k}{2\Delta x}, \quad \delta_y g_{i,j}^k = \frac{g_{i,j+1}^k - g_{i,j-1}^k}{2\Delta y}.$$

The finite difference operators for the second-order derivatives are,

$$\delta_x^2 g_{i,j}^k = \frac{g_{i+1,j}^k - 2g_{i,j}^k + g_{i-1,j}^k}{\Delta x^2}, \quad \delta_y^2 g_{i,j}^k = \frac{g_{i,j+1}^k - 2g_{i,j}^k + g_{i,j-1}^k}{\Delta y^2}.$$

At the boundary $j = 0$ an upwind discretization scheme is used [41] for the derivative in the y direction in (6.15) which reads,

$$\delta_y g_{i0}^k = \frac{-3g_{i0}^k + 4g_{i1}^k - g_{i2}^k}{2\Delta y}.$$

The mixed derivatives are discretized using the seven point stencil described in [49],

$$\delta_{xy} g_{ij}^k = \frac{1}{2\Delta x \Delta y}[g_{i+1,j+1}^k - 2g_{ij}^k + g_{i-1,j-1}^k] - \frac{\Delta x}{2\Delta y}[\frac{g_{i+1,j}^k - 2g_{i,j}^k + g_{i-1,j}^k}{\Delta x^2}] - \dots$$
$$\frac{\Delta y}{2\Delta x}[\frac{g_{i,j+1}^k - 2g_{i,j}^k + g_{i,j-1}^k}{\Delta y^2}].$$

We handle the Neumann boundary conditions (6.13) and (6.14) in the same way as described in [49]. The space discretization leads to a semi-discrete equation,

$$\frac{d\mathbf{g}}{dt} + \mathbf{A}\mathbf{g} = \mathbf{b},$$

where $\mathbf{A}$ is an $m(n+1) \times m(n+1)$ matrix and $\mathbf{b}$ is a column vector of length $m(n+1)$. The vector $\mathbf{g}$ of length $m(n+1)$ is the option price at the grid points. The vector $\mathbf{b}$ consists

of terms due to the Neumann boundary condition in the x direction and does not depend on t.

8.1.2 Time discretization

For time discretization we model similar to [62], the so called *Backward Difference Formula*, BDF2 scheme. This is an implicit scheme with second-order accuracy [62]. The stability of time discretization schemes is considered in [49] and [62] among others. At time $k\Delta t$ the BDF2 scheme reads,

$$\frac{3\mathbf{g}^{k+1} - 4\mathbf{g}^{k} + \mathbf{g}^{k-1}}{2\Delta t} + \mathbf{A}\mathbf{g}^{k+1} = \mathbf{b},$$

for $k = 1, 2, \ldots, l-1$. The favorable properties of the BDF2 scheme are considered in [62]. At each iterate of the BDF2 scheme we require the value of the last two iterates. As is typical, we obtain the first iterate using an Implicit Euler scheme [41]. That is, given $\mathbf{g}^0$ (the initial value), we obtain $\mathbf{g}^1$ using an implicit Euler scheme,

$$\frac{\mathbf{g}^{1} - \mathbf{g}^{0}}{\Delta t} + \mathbf{A}\mathbf{g}^{1} = \mathbf{b}.$$

We have verified numerically that this choice of space-time discretization of the initial/boundary-value problem gives us stable solutions for the option price at moderate grid sizes. At each time step we solve the following system of linear equation,

$$(\mathbf{I} + \frac{2}{3}\Delta t\mathbf{A})\mathbf{g}^{k+1} = \frac{4}{3}\mathbf{g}^{k} - \frac{1}{3}\mathbf{g}^{k-1} + \Delta t\mathbf{b}, \tag{8.1}$$

where $\mathbf{I}$ is an $m(n+1) \times m(n+1)$ identity matrix. We solve this system of equations using an LU decomposition.

After we obtain the option price at the discrete grid points, we solve the sensitivity

equations (7.12). The right-hand side of the sensitivity equations (7.12) is approximated using the central difference scheme for the first derivatives, the 3-point stencil for the second derivatives and the 7 point stencil for the mixed derivatives. We solve the sensitivity equations on the same grid that we used for the option price and use the same space-time discretization to discretize the sensitivity equations. We verify empirically that the scheme for the solution of the sensitivity equations converges.

8.2 Numerical study

8.2.1 Benchmark models

We present numerical results to study the impact of estimation error on option price for the following four Heston models from the 8 benchmark models presented in chapter 5. Note that Model_1 is the estimated Heston model corresponding to the 2006 S&P 500 data.

	κ	θ	γ	ρ
Model_1	16.6	.017	.28	-.54
Model_2	16.6	.017	.1	-.54
Model_3	25	.017	.28	-.54
Model_4	25	.017	.1	-.54

Table 8.1: Benchmark Heston models

For our study we will fix these four models as the true unknown models. We will fix the number of observations from which the estimators are observed to $N = 252$. We will obtain estimates for these four models using the estimation method described in chapter 3 given a set of $N = 252$ observations. The estimation error will be approximated by the square root of the diagonal terms of the matrix $L(P)/252$, where the asymptotic variance covariance matrix $L(P)$ for the 2006 S&P 500 parameters was obtained in chapter 5. We will then

obtain the matrix $L(P)$ corresponding to each model empirically by extensively simulating large number of trajectories each for reasonably large observation time. The estimation error for parameter P_i will then be approximated by $\sqrt{\frac{L(P)_{i,i}}{252}}$. To provide examples of practically usable sensitivity values, all displayed sensitivities are computed for $N = 252$.

For each of the four models in Table 8.1, we will study the impact of estimation errors on four distinct options. We pick options on the S&P 500 index for strike prices that were actively traded in the financial market during the first quarter of 2007.

	Strike Price : K	Time to maturity (days)
O_1	1380	63
O_2	1380	126
O_3	1440	63
O_4	1440	126

Table 8.2: Benchmark Option models

We have assumed in order to be consistent with practice, that the time step between daily observations is $T = 1/252$ so that the time to maturity T_M is equal to $T_M = .25$ for 63 days and $T_M = .5$ for 126 days.

8.2.2 The ranges of (SPX,VIX) data

To present our sensitivity results within a realistic range, we observe the SPX and VIX daily values for the first quarter of 2007. The SPX index values during this period are between 1370 and 1460. The corresponding VIX values are between 10 % and 20 %. We display the dependence of option price sensitivities on asset price and volatility within these two realistic ranges. In this period, the median SPX index value is 1426, and the median VIX index value is 11%. In order to facilitate a comparative study of the different options we

plot the option price for fixed volatility as a function of the ratio $s = \log(x/K)$ where x is the asset price and K is the strike price of the option.

8.2.3 Model identification

The unknown parameter vector P for each model in Table 8.1 is estimated using the constrained approximate maximum likelihood method described in chapter 3 from $N = 252$ observations. We present in Table 8.3 the estimated value Q for each of the four models estimated from $N = 252$ observations with time $T = 1/252$ between consecutive observations.

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$
Model_1	19	.02	.28	-.55
Model_2	14	.016	.09	-.44
Model_3	18.74	.018	.25	-.55
Model_4	25	.017	.09	-.53

Table 8.3: Estimated Heston models for $N = 252$ observations

The error in the estimator Q_i is computed by $\sqrt{\frac{L(P)_{i,i}}{N}}$ for $N = 252$ where $L(P)$ is the asymptotic variance-covariance matrix of the estimators and depends on the true parameter value P. The numerical implementation to obtain $L(P)$ was discussed in chapter 5. We simulate 1000 trajectories of the observations for the model associated to P, with N ranging from 500 to 15246. We compute the empirical variance-covariance matrix Σ_N over the 1000 trajectories. We observe from the numerical results of chapter 5 that for reasonably large N, $N\Sigma_N$ converges to a deterministic matrix. We take this limit as our estimate of the matrix $L(P)$. We observe that the off-diagonal entries of the matrix $L(P)$ are negligible. We report the value of the estimation error $\sqrt{L(P)_{i,i}/252}$ for each of the four models below in Table 8.4,

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$
Model_1	5.68	.002	.012	.06
Model_2	6	.001	.005	.06
Model_3	7	.002	.012	.06
Model_4	7	.001	.004	.06

Table 8.4: Estimation error, $\sqrt{L(P)_{i,i}/252}$

8.2.4 Estimation of λ

We obtain an estimate of λ using the method described in section 6.2.3. We pick an option traded actively between the 22 days trading period of Jan 03 2007 to Feb 02 2007 with strike price $K = 1425$ and maturity date Feb 17 2007. We observe the prices at which this option was traded during the period from Jan 03 2007 to Feb 02 2007 and compute the Black-Scholes implied volatility using the inbuilt Matlab function *impvbybls* for each day on which the trade occurred. In order to use simultaneous asset price and option price data we consider the closing asset price for each day, and the average of the closing option bid and ask price for each day. We then use the estimation formula (6.26) with parameter values estimated from the 2006 S&P 500 data to get a value of λ_t for each day during the observation period. We set the value of the risk free rate of return at $r = .01$ in (6.26). We compute the average of λ_t during this period of 22 days as our estimate of λ.

We validate this value by simultaneously comparing the mean squared error in time between predicted and actual option prices for a range of different λ values. We report concrete results here in Fig. 8.1 for one specific European option EOP_t with strike price 1430 and maturity date Feb 17 2007, observed daily from Jan 03 2007 to Feb 02 2007 . To compute the theoretical option price we used the estimated parameter values for the 2006 S&P 500 data set given in the first row of Table 8.1. Using the estimated $\lambda = 2$, we

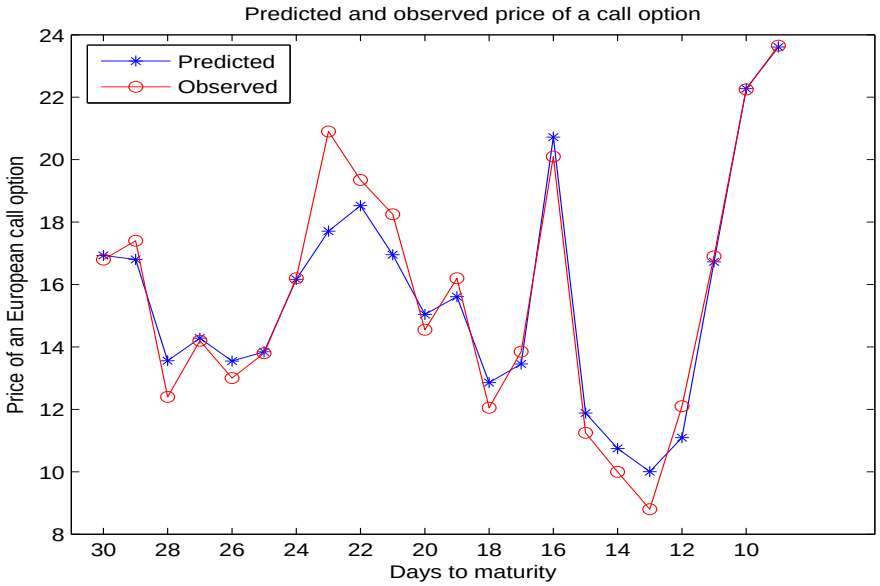


Figure 8.1: Predicted and observed option price of an European call option with strike price 1430 for $\lambda = 2$

solved the option price PDE (6.10) to compute dynamic estimates $E\hat{O}P_t$ of the option price EOP_t, and we derived the mean value 0.66 and standard deviation 0.96 of the prediction error $E\hat{O}P_t - EOP_t$ for the option price, which had a mean value of 15.63 during this period. Fig.8.1 represents the computed and market price of this European option EOP_t.

The results on the sensitivity of the option price with respect to λ are presented in section 8.2.7 for $\lambda = 2$. We conclude that for options with strike price close to the asset price (near the money options) the relative change in option price due to small changes in λ is less than 5% (see Table 8.6). Hence we chose and estimated a constant λ for this study. To study, in the next sections, how sensitive the option price is to the model parameters H and the impact of model estimation errors on the option price, we fix the value $\lambda = 2$ just estimated from market data.

8.2.5 Computational domain

Numerically we solve for the option price viewed at each time t as a function $h(lsp, y, t)$ of t, the log of the stock price ($lsp = \log(x)$), and of the squared volatility y of the stock price x. This change of variable gives a transformed initial/boundary value problem that we solve on the following computational domain for $[lsp, y, t]$, where

$$[lsp, y, t] \in (7, 8) \times (0, 1) \times (0, 0.25),$$

for O_1 and O_3 and

$$[lsp, y, t] \in (7, 8) \times (0, 1) \times (0, 0.5),$$

for O_2 and O_4. This domain was chosen after verifying carefully the robustness of the solution when the domain boundary $lsp = 7$ is reduced to $lsp = 5$. Since the time increment between two consecutive daily observations is set to 1/252 (by convention), the maturity date $T_M = .25$ indicates that we study an option with three months to maturity and $T_M = .5$ indicates that we study an option with six months to maturity. We set the risk free rate of return at value $r = .01$ and the market price of volatility risk λ at the estimated value $\lambda = 2$. We use the following grid size for our computation,

$$\Delta lsp = .016, \quad \Delta y = .006, \quad \Delta t = .004,$$

where Δlsp is the grid size for the space variable lsp computed as the "lsp" interval length (equal to 1 here) divided by the number of grid meshes in the lsp direction. Table 8.5 displays the CPU-times in seconds on a standard PC required for solving the partial differential equations for the option price PDE and all 5 sensitivity PDEs. The LU decomposition is performed only once for each solution. The key computational cost is the LU decomposition

Grid size $(m \times n \times s)$	Δl	Δy	Δt	CPU-time (secs)
226800	1/60	1/60	1/252	95
308700	1/70	1/70	1/252	240

Table 8.5: CPU time for option price & derivatives

and backward substitution for a matrix of size $m(n+1) \times m(n+1)$ where m and n are the grid sizes in the lsp and the y direction. The CPU-times increase more rapidly as we increase the spatial grid size $m(n+1)$ than when we increase the grid size in time. For coarser grids the CPU-times are much lower but the solution is not accurate. We performed extensive numerical simulations to verify that the solutions converge and that the boundary conditions are satisfied.

8.2.6 Impact of parameter estimation errors on option pricing

Using the method described in section 7.2 we compute the impact of parameter estimation errors on the option price through the formula,

$$\varepsilon(Q) = \max_{P \in B(Q)} \delta(P),$$

for each Q in Table 8.3 where the L^2-error $\delta(P)$ is computed by Equation(7.6). The neighborhood $B(Q)$ is the product of the 4 intervals $B_i = [Q_i - \sigma_i(P), Q_i + \sigma_i(P)],\ i = 1, 2, 3, 4$ where

$$\sigma_1(P) = \sqrt{\frac{L(P)_{1,1}}{252}}, \quad \sigma_2(P) = \sqrt{\frac{L(P)_{2,2}}{252}}, \quad \sigma_3(P) = \sqrt{\frac{L(P)_{3,3}}{252}}, \quad \sigma_4(P) = \sqrt{\frac{L(P)_{4,4}}{252}}$$

are the estimates for standard deviation of κ, θ, γ and ρ corresponding to the model with true parameters P.

We discretize the set $B(Q)$ using three grid points for each parameter, which gives us a

grid of $3^4 = 81$ vectors. At each one of these 81 grid vectors P we compute the four partial derivatives of the option price with respect to the four key model parameters, in order to compute the L^2-error $\delta(P)$. We then take the maximum of these 81 values of $\delta(P)$ to obtain the option pricing error $\varepsilon(Q)$.

We first plot the price of the four benchmark options under the four estimated models in Table 8.3. The graphs in Fig. 8.2 show the price of the options O_1, O_2 and the graphs in Fig. 8.3 show the price of the options O_3 and O_4 respectively. In each graph we plot the option price as a function of the asset price for volatility fixed at 11%. The abscissa is the ratio $s = \log(x/K)$ where x is the asset price and K the strike price of the option. The option price is displayed for s between $-.05$ to .05, which corresponds to x between 1310 to 1450 for O_1 and O_2 and x between 1360 and 1510 for O_3 and O_4.

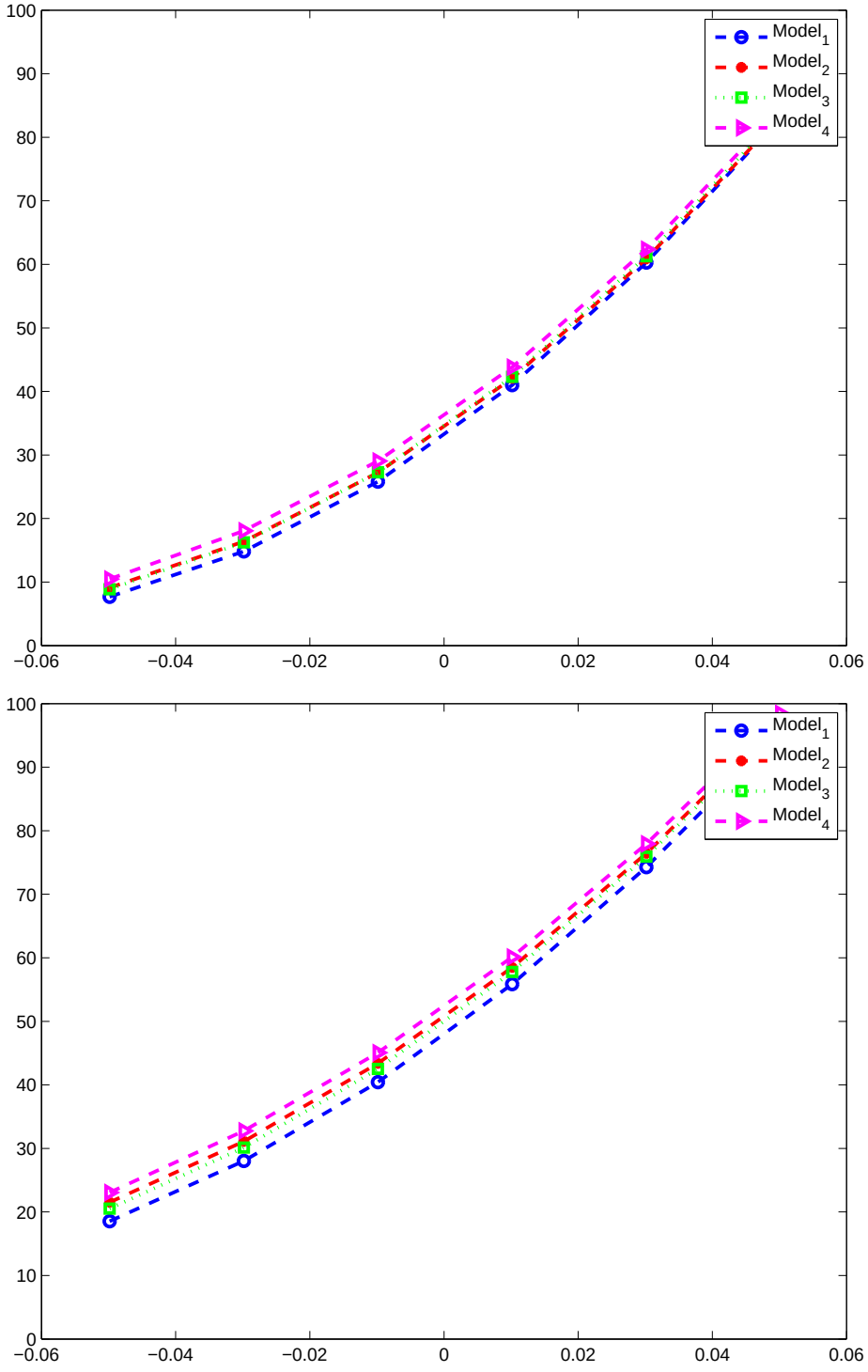


Figure 8.2: Price of O_1(top) and O_2(bottom) for the four Heston models for volatility = 11%

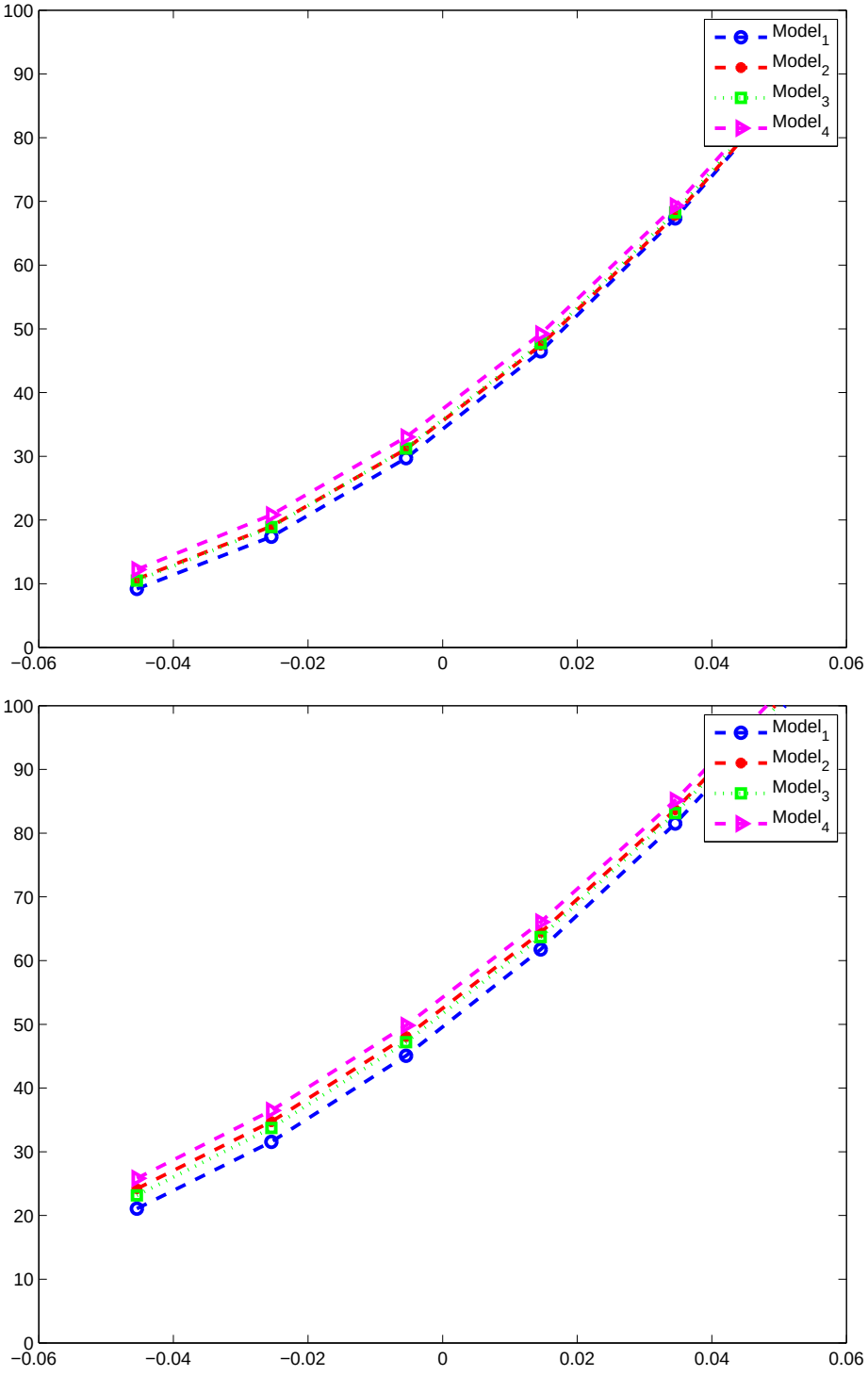


Figure 8.3: Price of O_3(top) and O_4(bottom) for the four Heston models for volatility = 11%

In Fig. 8.4 and 8.5 we illustrate the impact of parameter estimation errors on the pricing of the four options respectively from Table 8.2. In each figure we plot the option pricing errors for the four underlying Heston models in Table 8.1. We plot for each case the error in the option price due to estimation error defined by the bound $\varepsilon(Q)$. In Fig. 8.4 and Fig. 8.5 the errors are plotted as a function of the asset price. The abscissa is the ratio $s = \log(x/K)$ where x is the asset price and K is the strike price of the option. We present the results on an extended domain in order to see the impact of estimation errors on deep-out-of-the money options (when strike price is much larger than the trading price of the underlying asset) and deep-in-the money options (when strike price is much smaller than the trading price of the underlying asset). The volatility value is fixed at 11%. The value of s for which the estimation error results are displayed are between $s = -.11$ to $s = .11$ which corresponds to asset price value x between 1230 and 1540 for O_1 and O_2 and x between 1280 and 1600 for O_3 and O_4.

We observe that the sensitivity is highest when the asset price is close to the strike price of the option and decreases as the asset price goes away from the option strike. In other words options that are close to the money are more sensitive to estimation errors than options that are far from the money. We note here that for the Black-Scholes model, the derivative of the option price with respect to the volatility parameter is indeed highest when the asset price is equal to the strike price and decreases as the asset price goes away from the strike price.

We also see by comparing the results between the four option models that options which are farther from maturity are more sensitive to estimation errors as compared to options which are closer to maturity. We observe that estimation impact is larger on Model_1 and Model_3 with values between 2 to 5 for Model_1 and 1.5 to 3 for Model_3. We note that the volatility γ of the squared volatility process is .28 for these two models and .1 for Model_2

and Model_4.

In Fig. 8.6 and Fig. 8.7 we display the option price together with the error $\varepsilon(Q)$ added to and subtracted from it. We display the option price for realistic values of asset price for all four options for Model_1 with volatility fixed at 11%. The solid blue line displays the option price, while the two dotted red lines are the option price plus/minus the error bound $\varepsilon(Q)$. The abscissa is the ratio $s = \log(x/K)$ where x is the asset price and K the strike price of the option. The option price is displayed for s between $-.05$ to .05, which corresponds to x between 1310 to 1450 for O_1 and O_2 and x between 1360 and 1510 for O_3 and O_4.

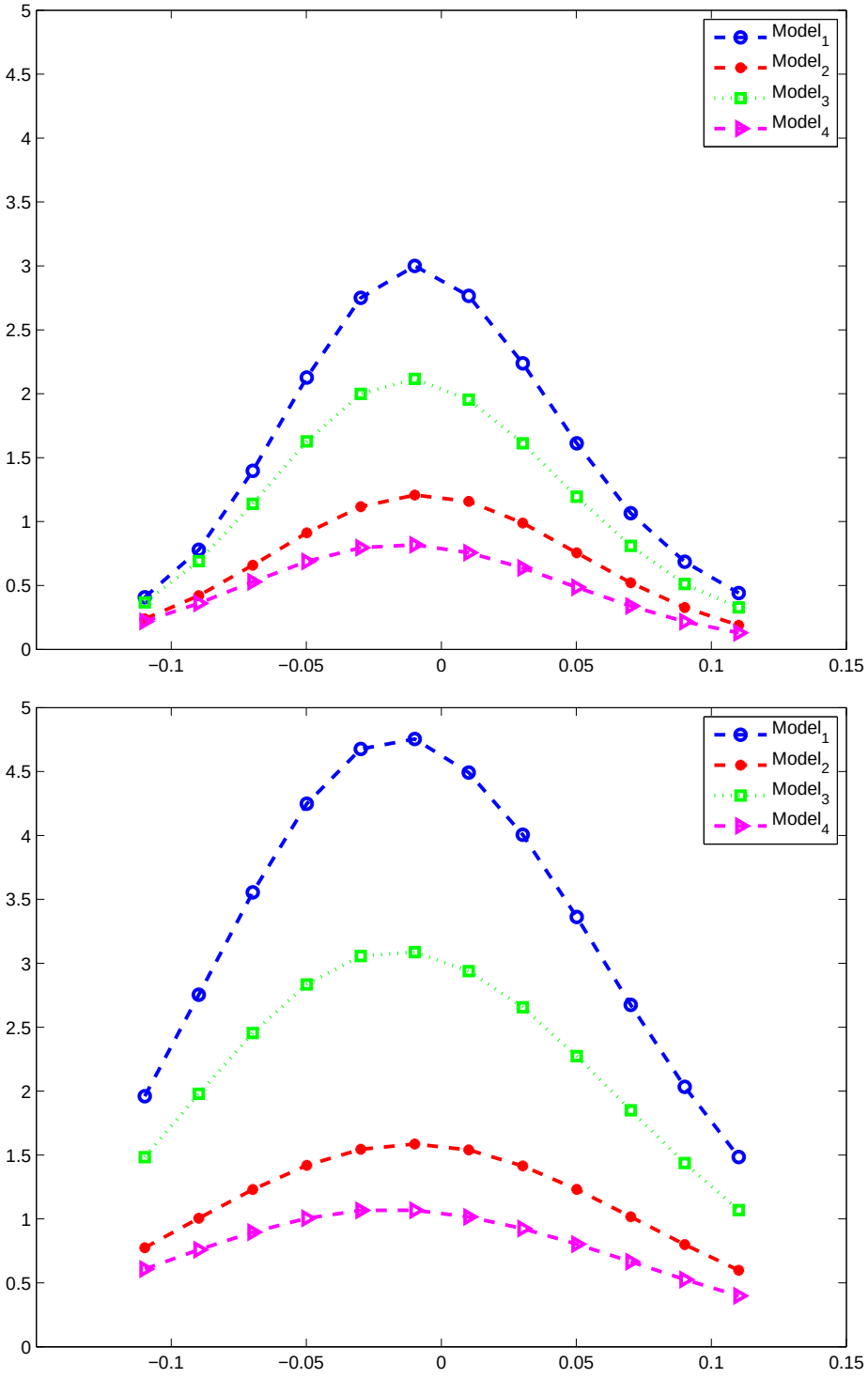


Figure 8.4: Estimation error impact on O_1(top) and O_2(bottom) for the four Heston models for volatility = 11%

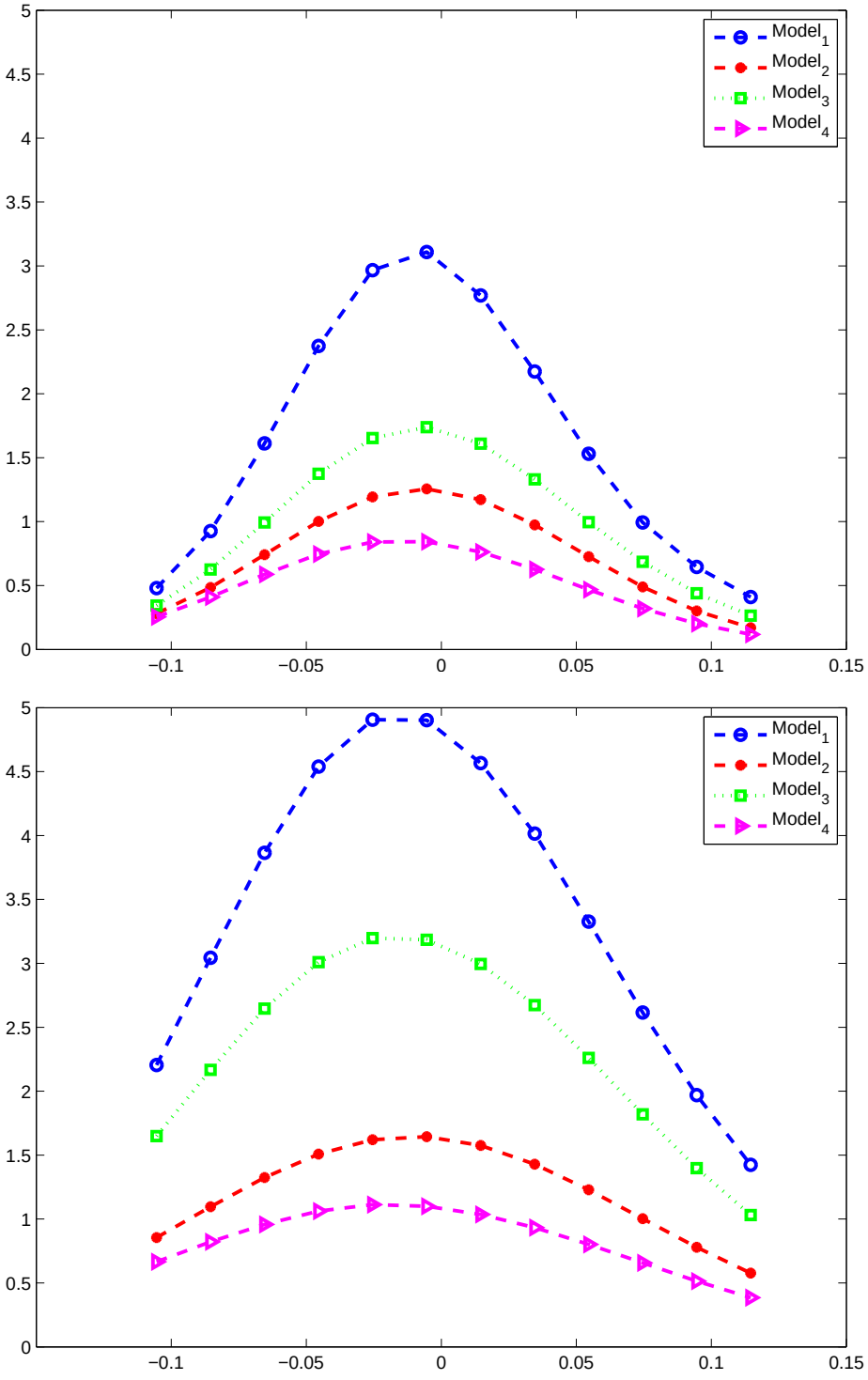


Figure 8.5: Estimation error impact on O_3(top) and O_4(bottom) for the four Heston models for volatility = 11%

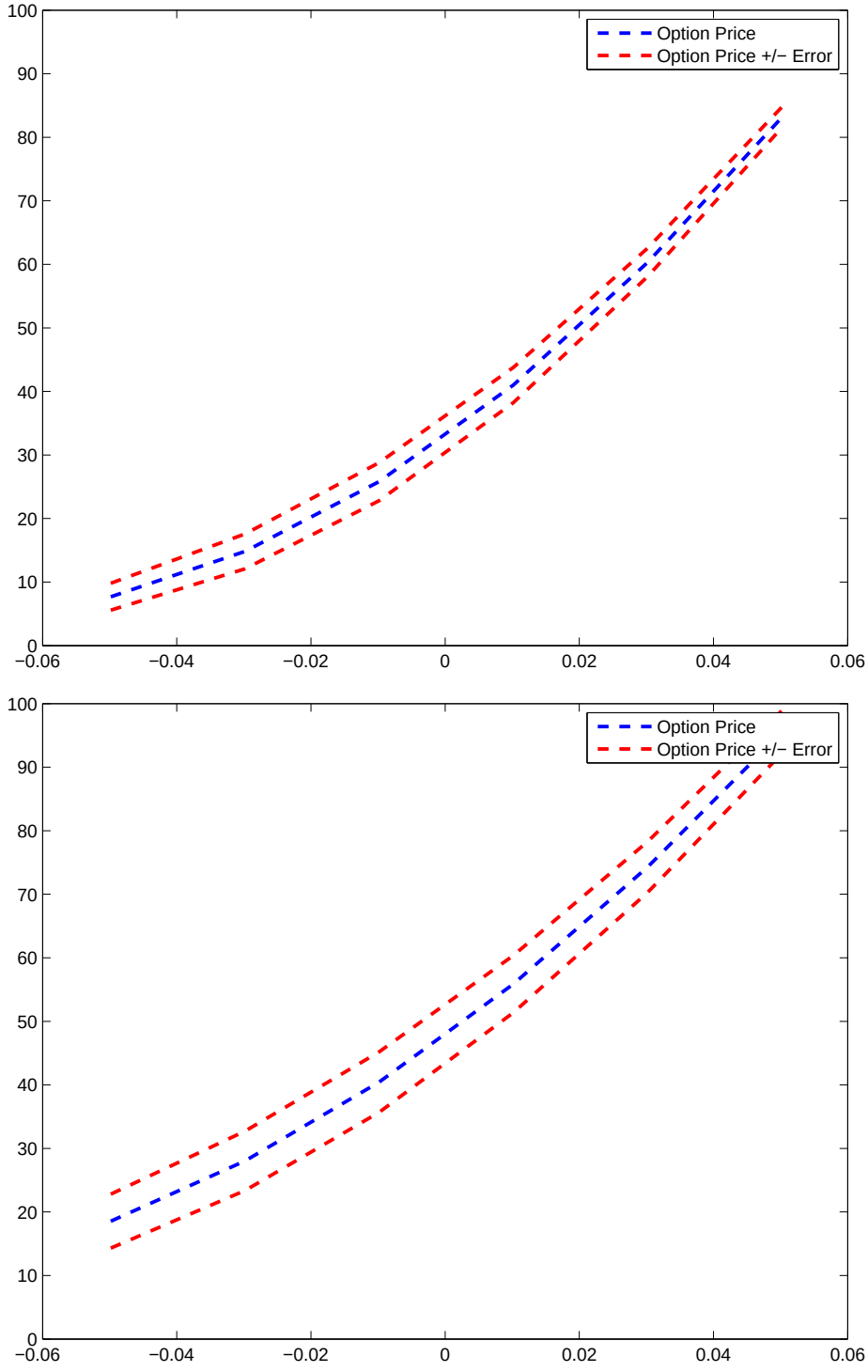


Figure 8.6: Option price +/- $\varepsilon(Q)$ under Model$_1$ for O_1(top) and O_2(bottom) at volatility = 11%

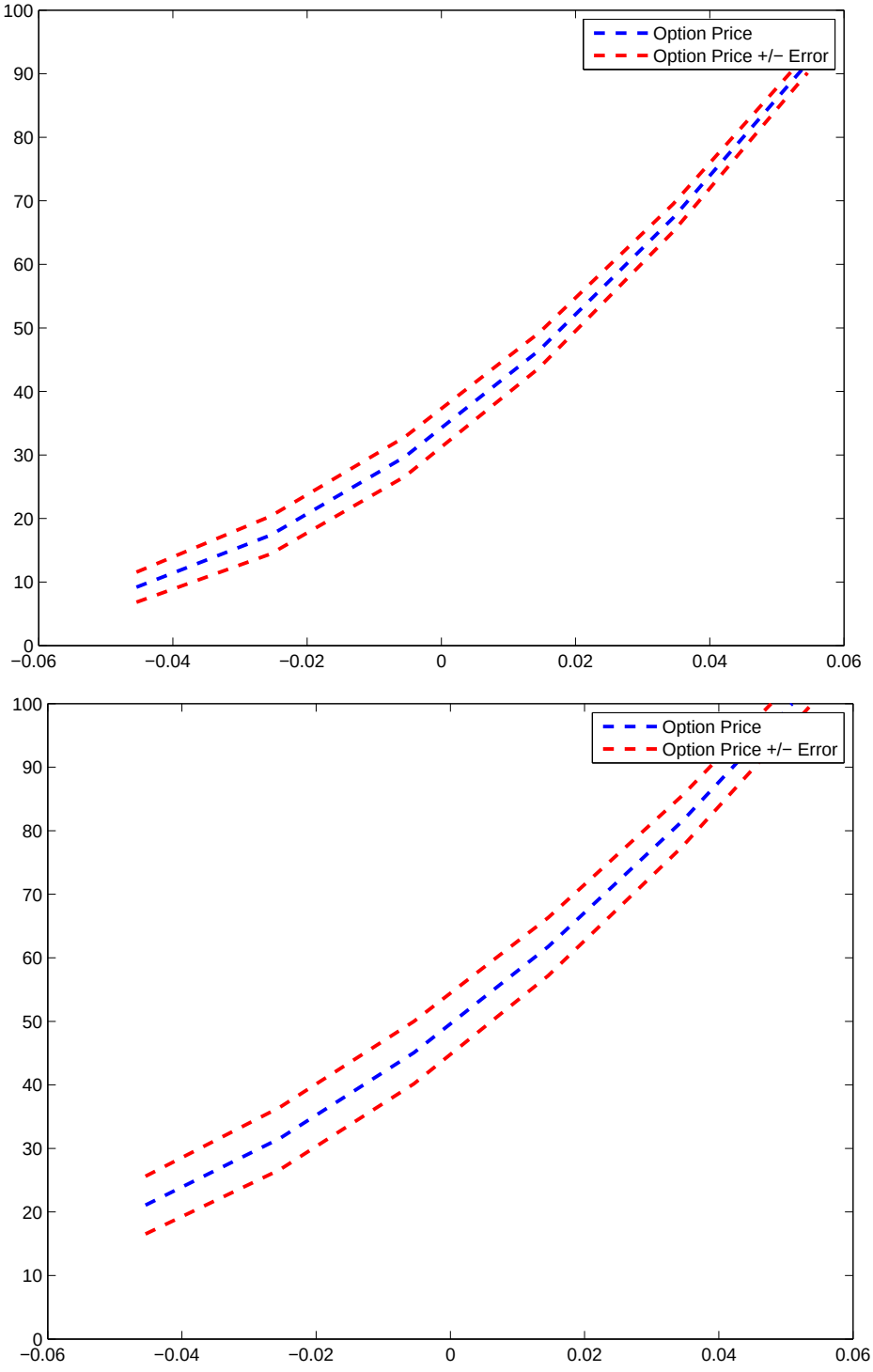
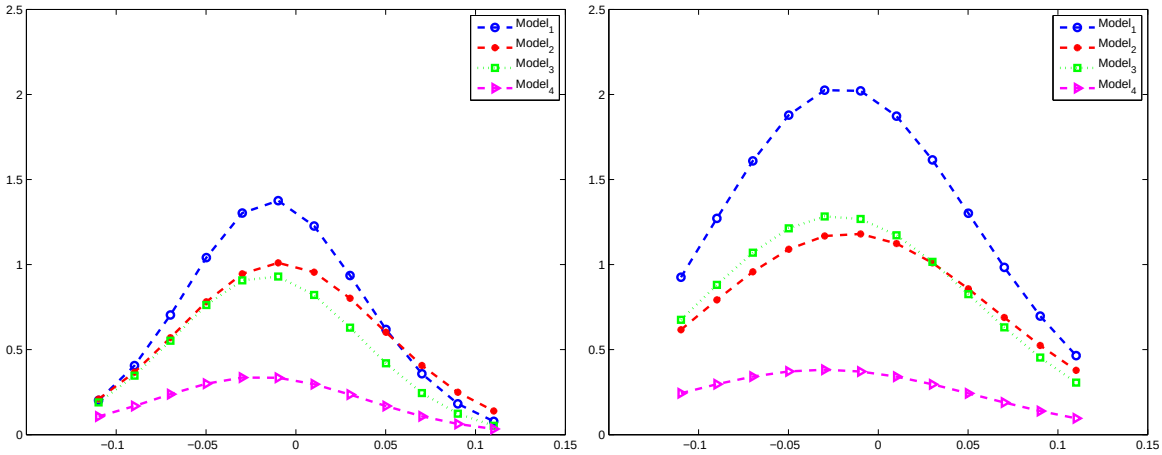


Figure 8.7: Option price +/- $\varepsilon(Q)$ under Model$_1$ for O_3(top) and O_4(bottom) for volatility = 11%

We next plot the individual sensitivity of the option price to each model parameter namely κ, θ, γ and ρ. Let $H = (\kappa, \theta, \gamma, \rho)$. Recall that the sensitivity with respect to parameter H_i, $i = 1, 2, 3, 4$, corresponding to a model with true parameter vector P is estimated by $|D_{i,Q}|\sqrt{\sigma_{i,i}(Q)}/g(Q)$ where Q is an estimate of P, $D_{i,Q}$ is the partial derivative of the option price $g(Q)$ with respect to parameter H_i, and $\sqrt{\sigma_{i,i}(Q)}$ is $\sqrt{L(P)_{i,i}/252}$.

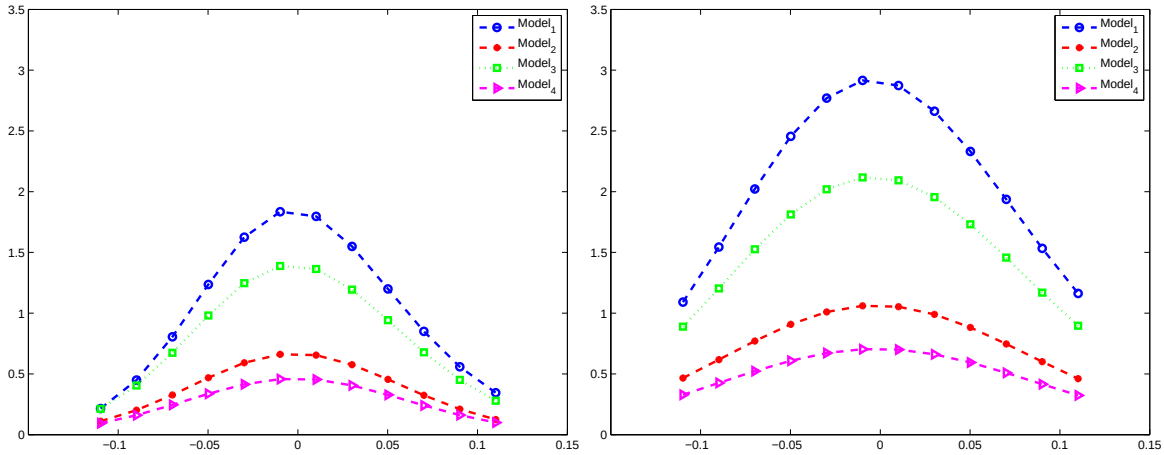
In Fig. 8.8, 8.9, 8.10, and 8.11 we illustrate the option price sensitivity with respect to parameters κ, θ, γ and ρ respectively. In each figure we plot four subfigures for the four benchmark option models. Each subfigure displays for a fixed option the sensitivity with respect to a fixed parameter for the four Heston models. We plot the sensitivities as a function of the asset price with volatility fixed at 11%. The abscissa is $s = \log(x/K)$ where x is the asset price and K is the strike price of the option. The value of s for which the estimation error results are displayed are between $s = -.11$ to $s = .11$ which corresponds to asset price value x between 1230 and 1540 for O_1 and O_2 and x between 1280 and 1600 for O_3 and O_4. We observe that the sensitivities with respect to κ and θ are higher as compared to γ and ρ.



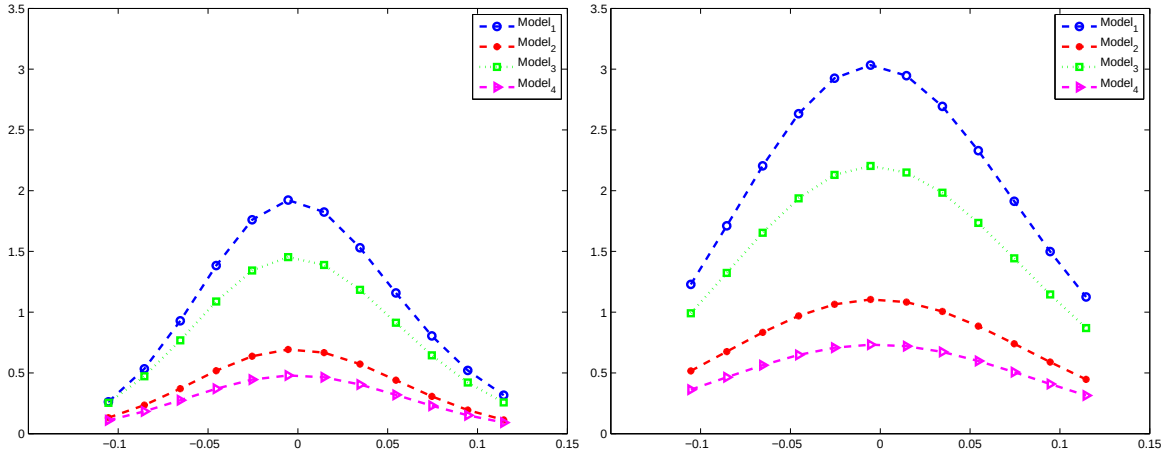
(a) O_1 (b) O_2

(c) O_3 (d) O_4

Figure 8.8: Sensitivity of the option price with respect to κ for volatility = 11%

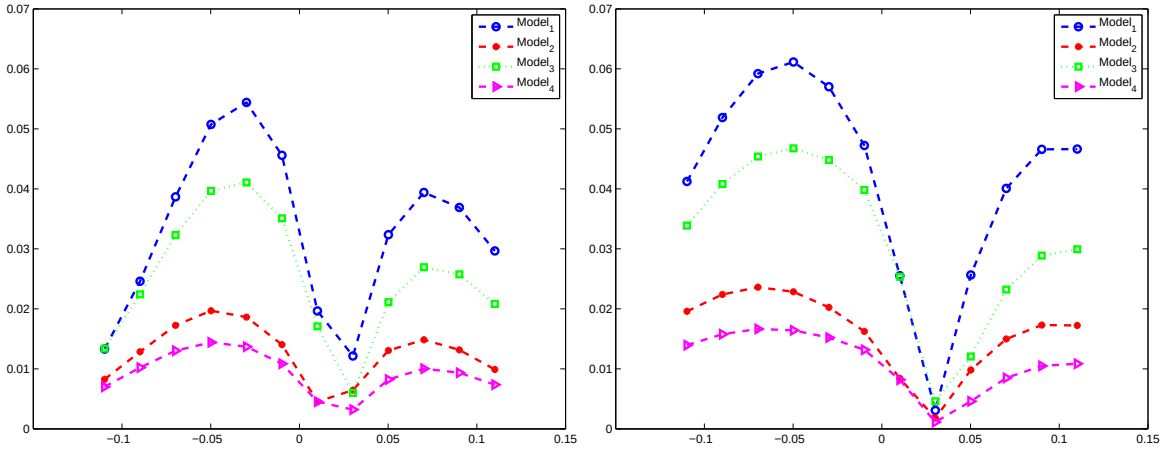


(a) O_1 (b) O_2



(c) O_3 (d) O_4

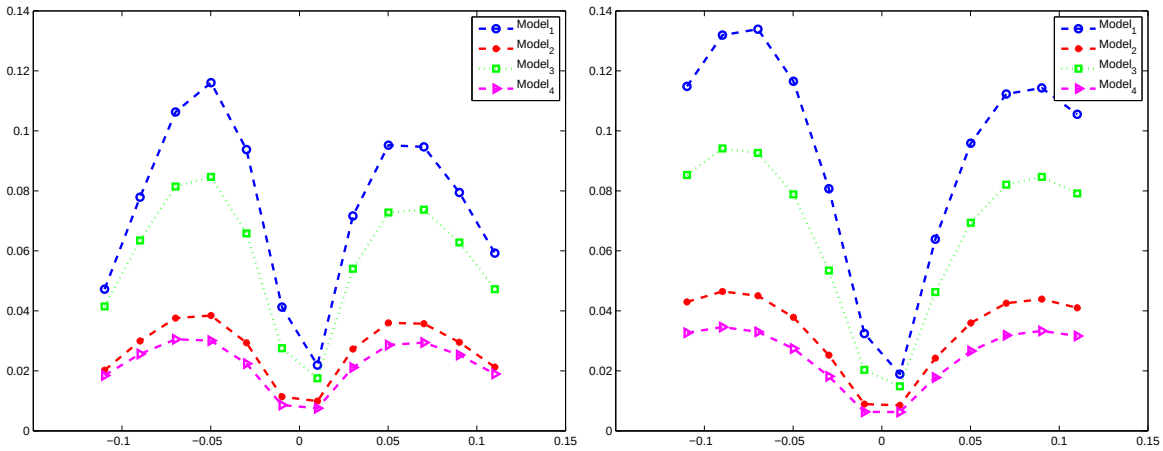
Figure 8.9: Sensitivity of the option price with respect to θ for volatility = 11%



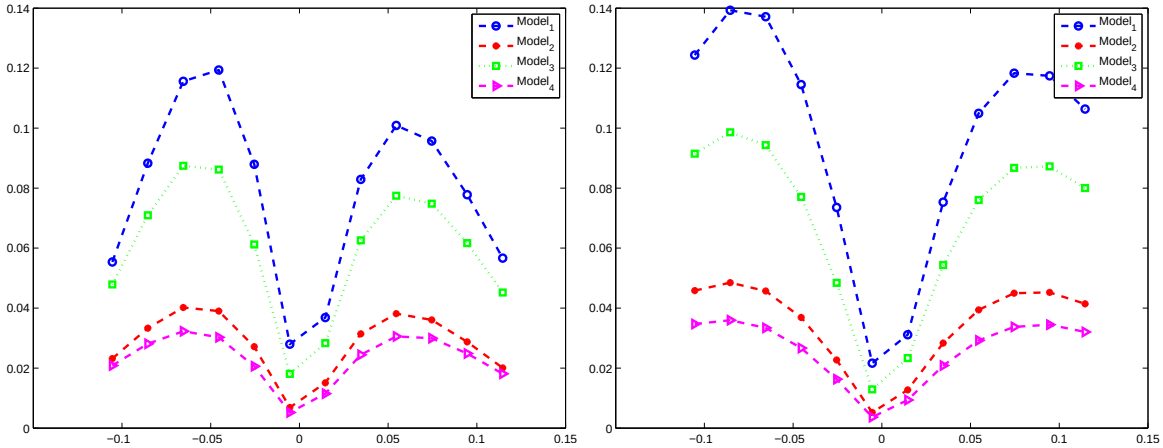
(a) O_1 (b) O_2

(c) O_3 (d) O_4

Figure 8.10: Sensitivity of the option price with respect to γ for volatility = 11%



(a) O_1 (b) O_2



(c) O_3 (d) O_4

Figure 8.11: Sensitivity of the option price with respect to ρ for volatility = 11%

We now plot for each case the *relative pricing error* $\varepsilon(Q)/g(Q)$, where the bound $\varepsilon(Q)$ on the option pricing error is divided by the option price $g(Q)$. In Fig. 8.12 and Fig. 8.13 the relative errors are plotted as a function of the asset price where the abscissa is the ratio $s = \log(x/K)$ where x is the asset price. The volatility value is fixed at 11%. The value of s for which the estimation error results are displayed are between $s = -.11$ to $s = .11$ which corresponds to asset price value x between 1230 and 1540 for O_1 and O_2 and x between 1280 and 1600 for O_3 and O_4. The individual relative sensitivities are plotted in Fig.8.14, Fig. 8.15, Fig. 8.16 and 8.17. Since the call option price increases with s, the decrease in relative errors as s increases does not say much. However we see that for at the money options, i.e, when the the ratio $s = x/K$ is close to zero, the relative sensitivity of the option price to θ (Fig. 8.15) is between 5% to 10% while the sensitivity to κ (Fig. 8.14) is between 1% to 5%. The overall impact of estimation errors for close to the money options is between 8% to 12% for the four benchmark options under the four models (Fig. 8.12-8.13).

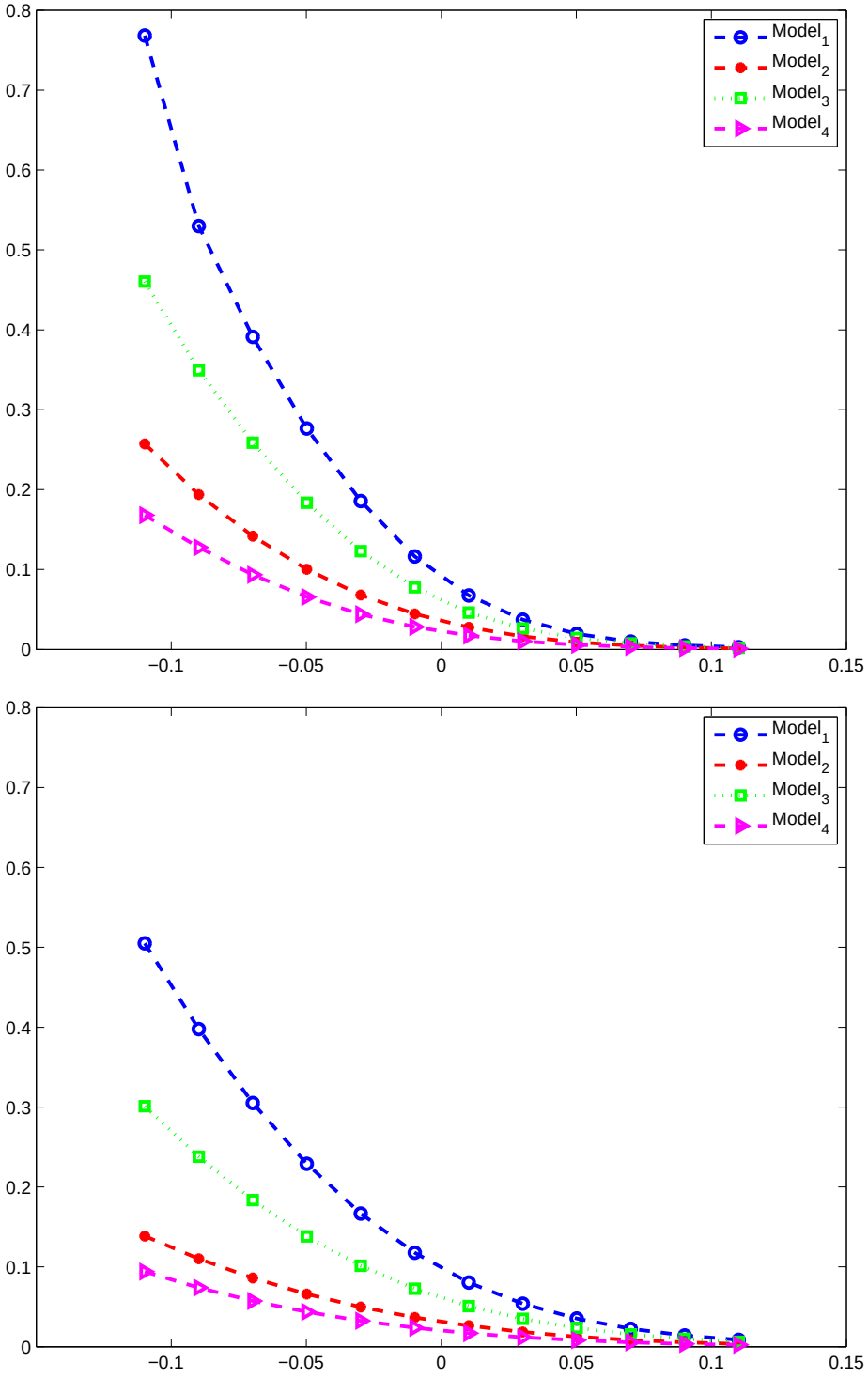


Figure 8.12: Relative estimation error impact on O_1(top) and O_2(bottom) for the four Heston models for volatility = 11%

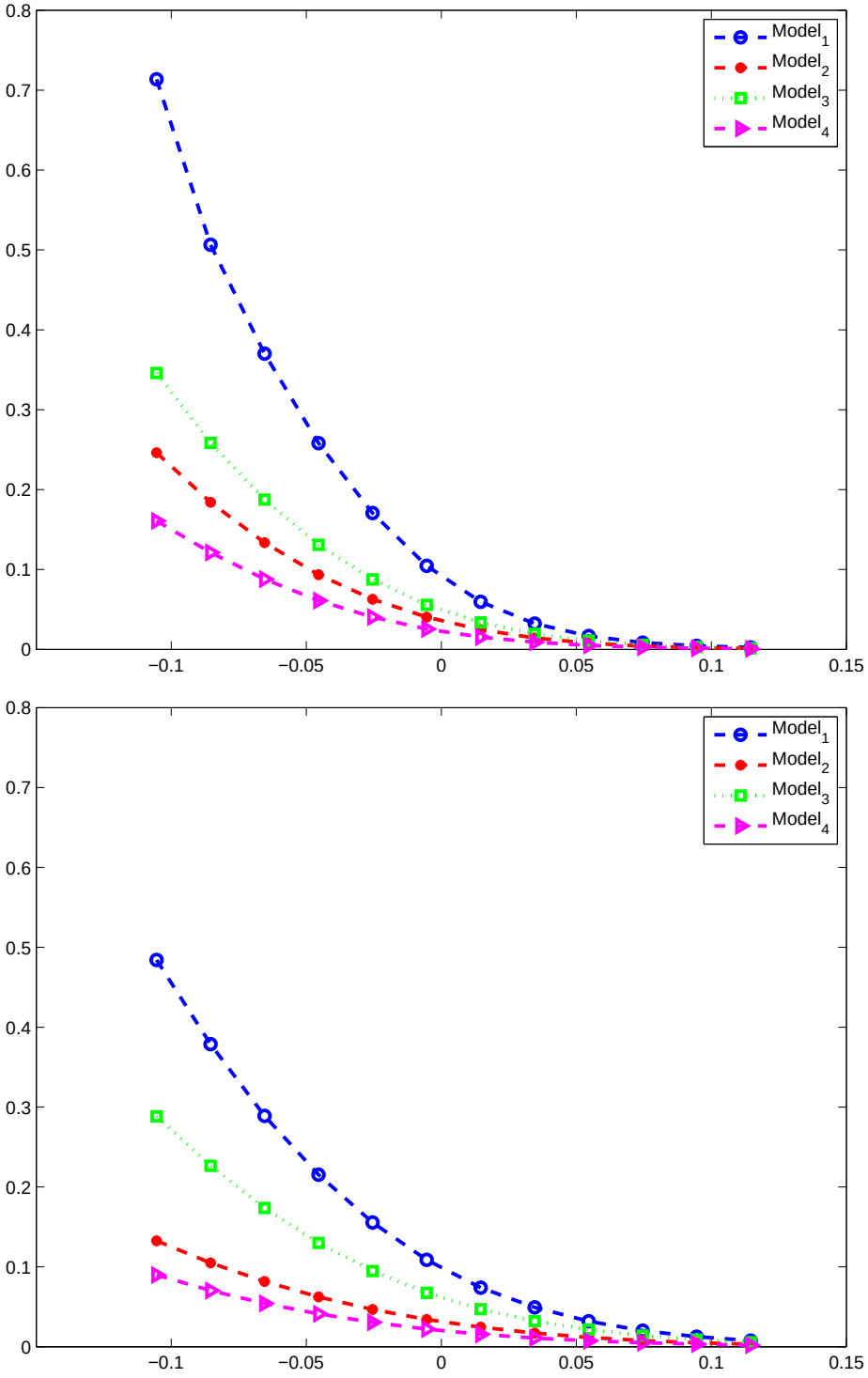
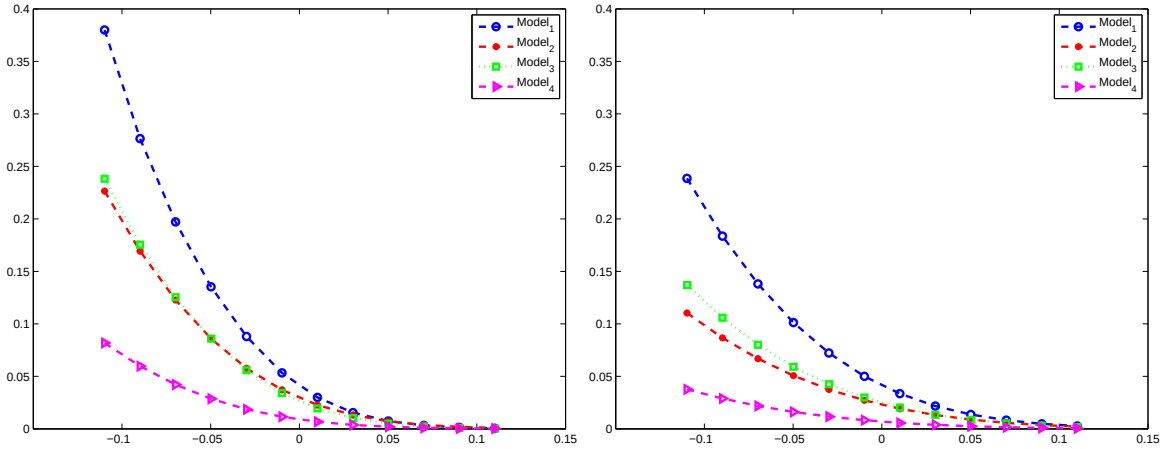
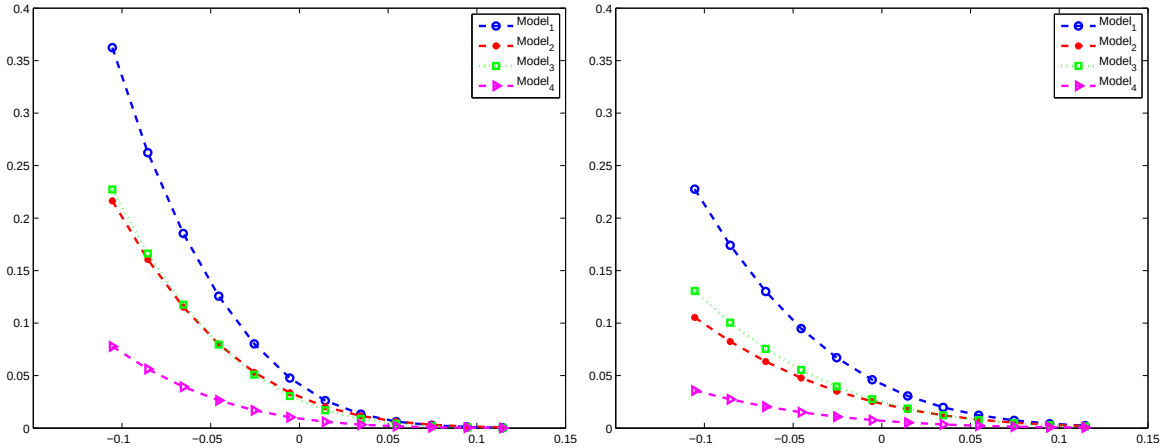


Figure 8.13: Relative estimation error impact on O_3(top) and O_4(bottom) for the four Heston models for volatility = 11%

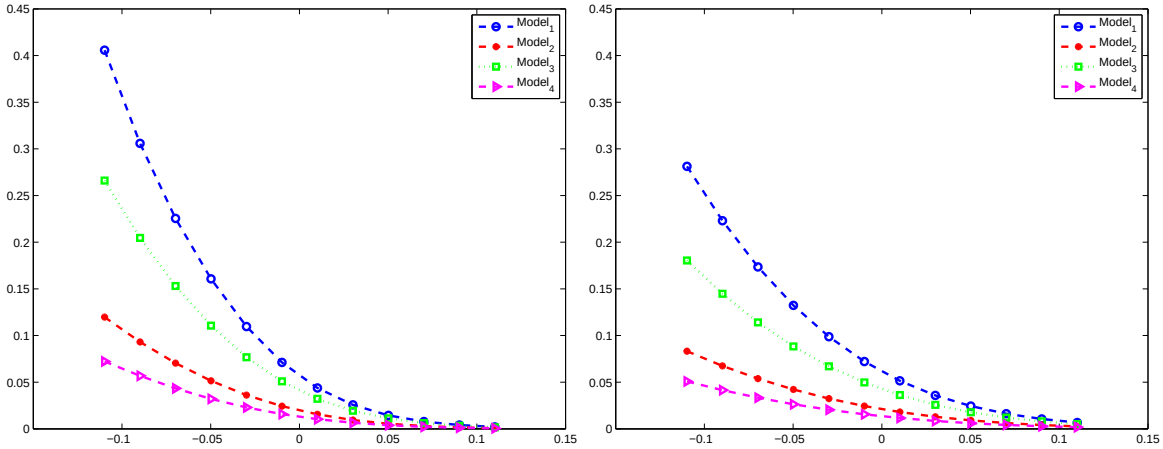


(a) O_1 (b) O_2

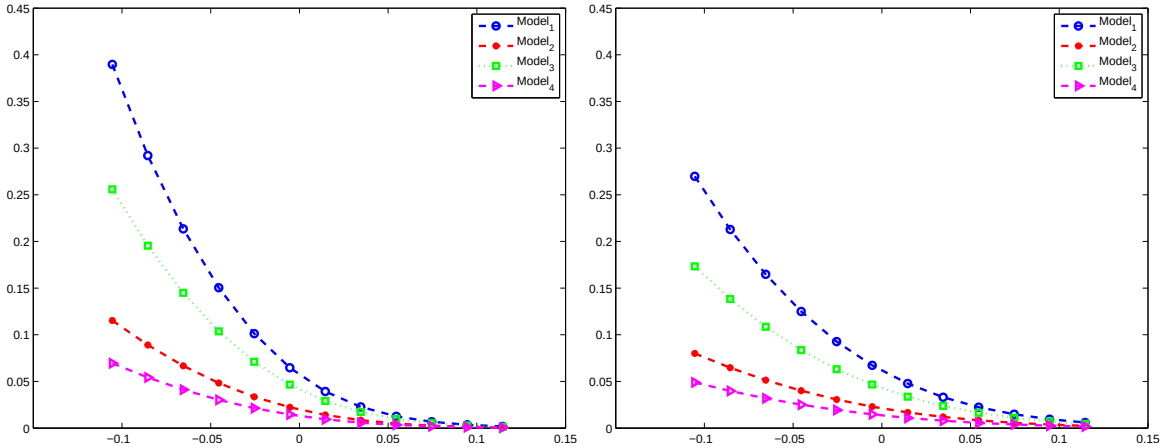


(c) O_3 (d) O_4

Figure 8.14: Relative sensitivity of option price with respect to κ for volatility = 11%

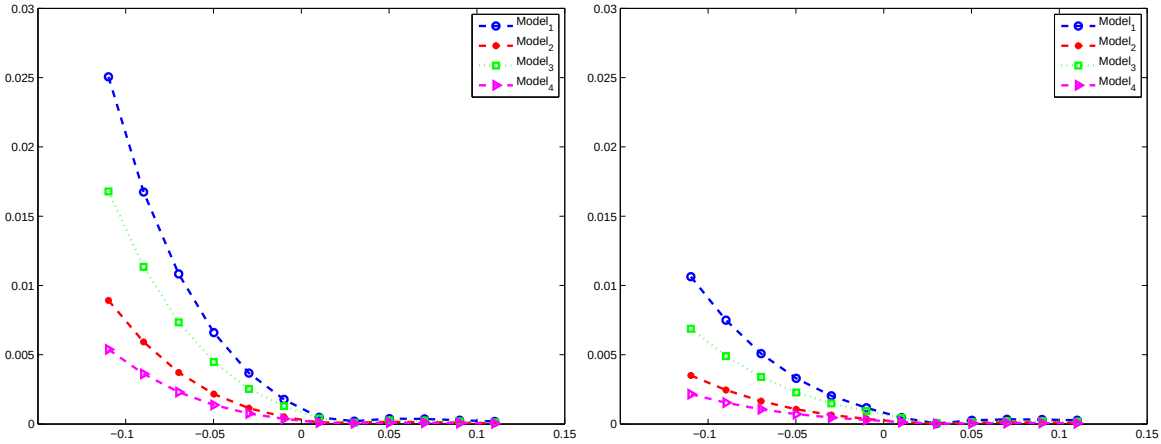


(a) O_1 (b) O_2



(c) O_3 (d) O_4

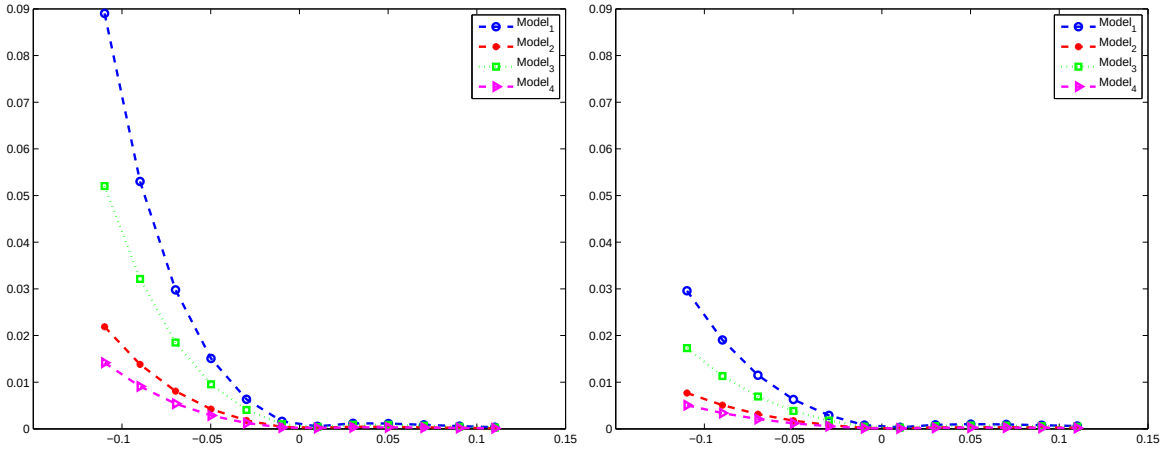
Figure 8.15: Relative sensitivity of option price with respect to θ for volatility = 11%



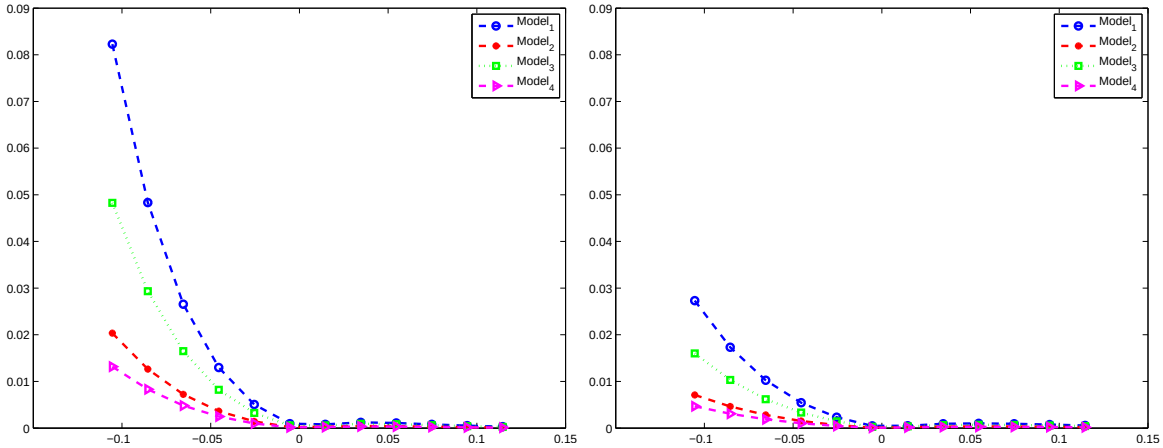
(a) O_1 (b) O_2

(c) O_3 (d) O_4

Figure 8.16: Relative sensitivity of option price with respect to γ for volatility = 11%



(a) O_1 (b) O_2



(c) O_3 (d) O_4

Figure 8.17: Relative sensitivity of option price with respect to ρ for volatility = 11%

Finally in Fig. 8.18, 8.19, 8.20, and 8.21 we plot the option price as a function of the volatility variable with asset price fixed at 1360 for O_1 and O_2 and asset price fixed at 1420 for O_3 and O_4. The volatility range are for which the results are displayed is from 11% to 22%. We plot the option price and impact of estimation error for the four options and the four underlying models. The left graph in each figure is the option price and the graph on the right is a display of the option price error as a function of volatility. Clearly, the option price increases as a function of volatility. The option price errors seem to decrease and then increase as the volatility increases. We see that the option price under Model_1 has the highest impact of estimation errors as a function of volatility.

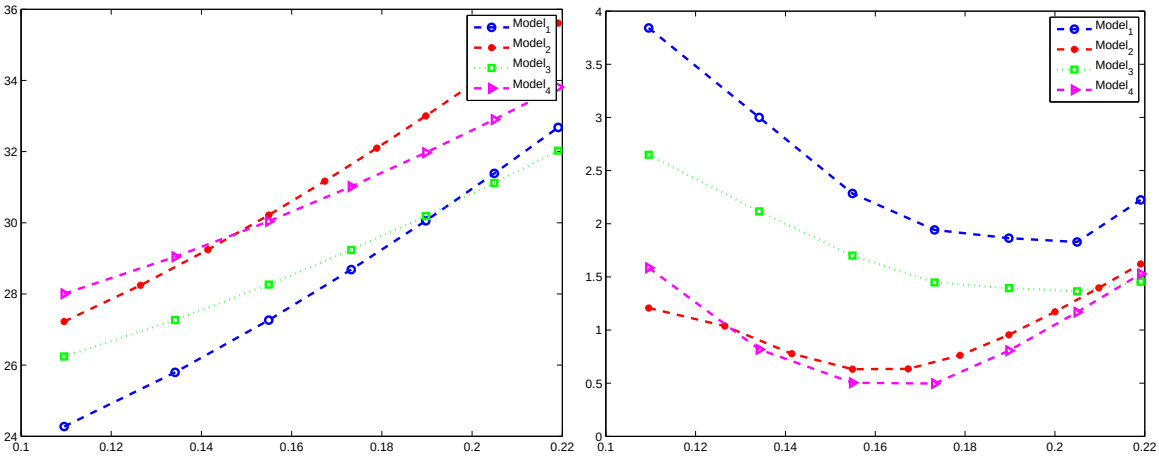


Figure 8.18: Price (left) and impact of estimation error (right) on O_1 as a function of volatility for asset price = 1360

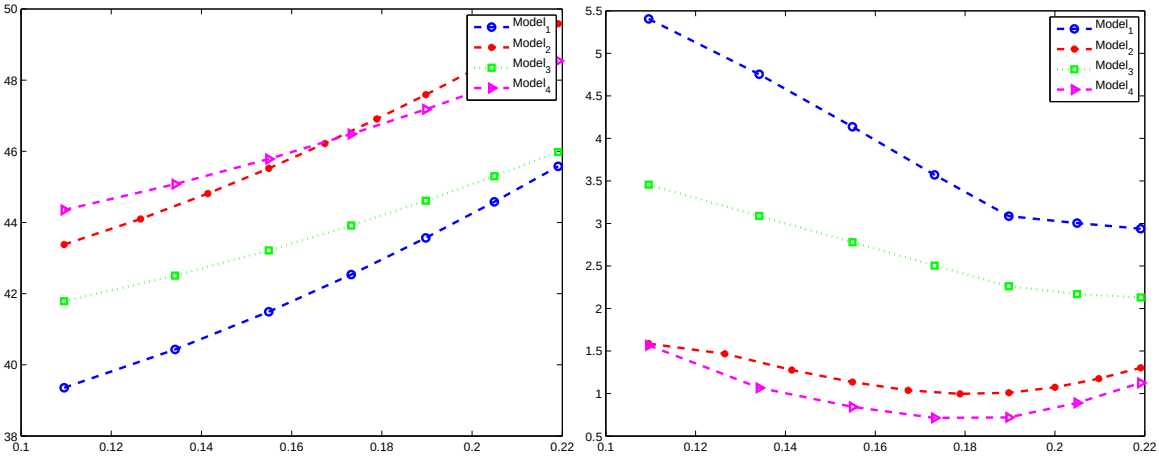


Figure 8.19: Price (left) and impact of estimation error (right) on O_2 as a function of volatility for asset price = 1360

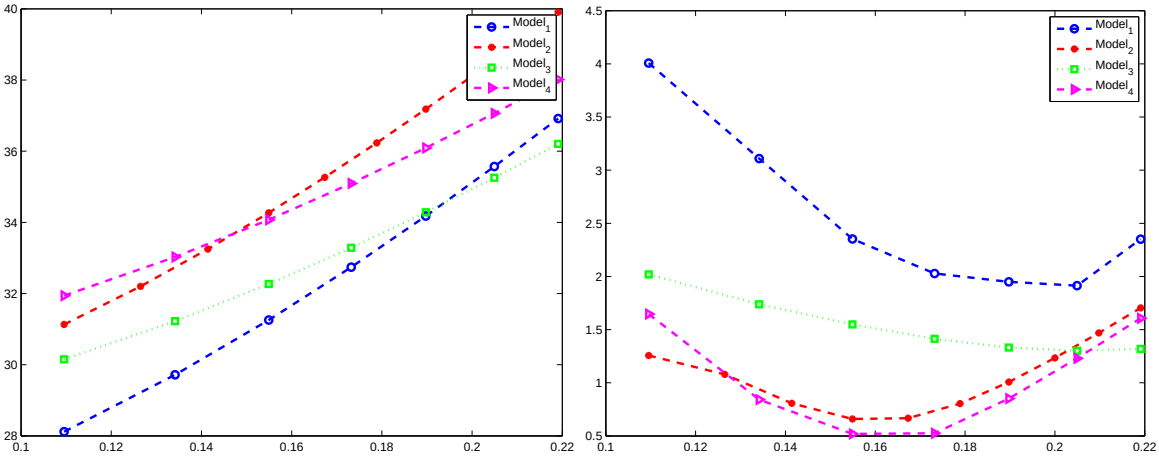


Figure 8.20: Price (left) and impact of estimation error (right) on O_3 as a function of volatility for asset price = 1420

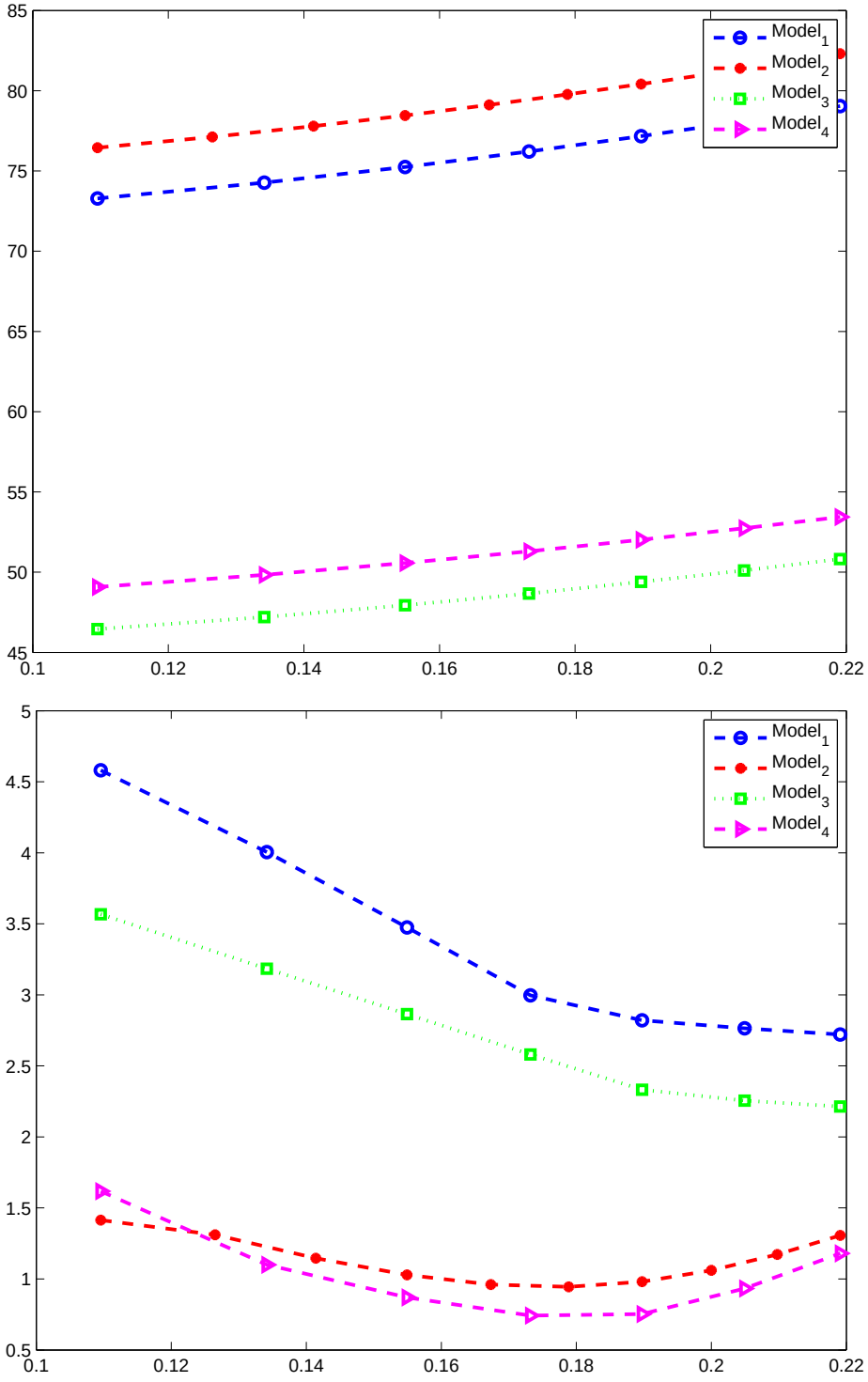


Figure 8.21: Price (top) and impact of estimation error (bottom) on O_4 as a function of volatility for asset price = 1420

Asset price	Volatility	Derivative	Option price
1422	11%	1.55	29.90
1422	19%	1.64	35.74
1357	11%	.96	5.65
1357	19%	1.07	9.29
1480	11%	1.30	67.43
1480	19%	1.45	72.35

Table 8.6: Absolute value of partial derivative of the option price with respect to λ computed at $\lambda = 2$ for an SPX option with strike price 1430 and 3 months to maturity

8.2.7 Option price sensitivity to market price of volatility risk

We compute the derivative $D_5(g(Q)) = \partial_\lambda g(Q)$ of the option price with respect to the market price of risk, to evaluate the option price sensitivity to errors on this unknown market price λ. Table 8.6 presents $|\partial_\lambda g(Q)|$ at realistic levels of asset price and volatility values. These derivatives are computed at the estimated model Q for an option with strike price equal to 1430 and 3 months to maturity.

8.3 Conclusion

We have developed a method which combines consistent estimation of the underlying Heston model parameters and of their variances and covariances, with the numerical solution of six parabolic partial differential equations in $\mathbb{R}^2$, in order to compute the impact of the unavoidable estimation errors on the price of options.

We have shown that the estimators from our approximate maximum likelihood approach are consistent and all available in closed forms. We have validated numerically the consistency of the estimators and studied the small sample properties of the estimators for the 2006 S&P 500 daily data. For the computation of option price under the Heston model we have developed and applied an optimal method to estimate the market price of volatility risk from concrete options data.

We have applied our approach to model the 2006 S&P 500 daily data by a Heston pair of coupled SDEs, and to study the option price sensitivities of several European call options based on this index. We have further applied our method to do a comparative study of a group of eight benchmark Heston models and four option models each time for practical values of N, the number of observations and T, the time between observations. The results are coherent from a pragmatic point of view.

Chapter 9

Technical annex - Joint parameter estimation for Heston model

We compute the estimators κ, θ, γ, and ρ of the Heston model parameters from the *approximate joint density* of the two processes driving the Heston model and compare the accuracy of these four estimators obtained from joint estimation to the accuracy of estimators from the decoupled estimation method discussed in the chapter 3. The joint estimation requires numerical optimization because the estimators obtained from solving first-order optimality conditions cannot all be expressed explicitly in terms of the data only. We compare the computation times of the two estimation procedures in section 9.2.

9.1 Approximate log-likelihood based on Euler discretization of joint model

We recall the joint dynamics under the Heston model of the asset price, X_t and the square of asset price volatility, Y_t,

$$dX_t = \mu X_t dt + \sqrt{Y_t} X_t dZ_t \tag{9.1}$$

$$dY_t = \kappa(\theta - Y_t)dt + \gamma\sqrt{Y_t} dB_t \tag{9.2}$$

where the processes Z and B are standard Brownian motions with $E[dZ_t dB_t] = \rho dt$. The joint density of the coupled SDEs for asset price and volatility is not known in closed form. Similarly to our approach for the decoupled estimation method we discretize the Heston model using the classical Euler [52] and then compute maximum likelihood estimators of the discretized model. The convergence of the Euler discretization scheme for the Heston model was discussed in chapter 3.

Let $H = (\kappa, \theta, \gamma, \rho, \mu)$ denote the vector of model parameters. For the joint estimation we will study only Case 1 described in the previous chapter because If Case 1 is not applicable to our data, then we do not accept the estimated parameters, since a priori the model parameters satisfy the strict constraints described below,

$$\kappa > 0,\, \theta > 0,\, \gamma > 0, \quad 0 < \theta < 1, \quad -1 \leq \rho \leq 1, \quad 2\kappa\theta > \gamma^2.$$

Define the feasible domain $\mathcal{C}_H \subset \mathbb{R}^5$ by,

$$\mathcal{C}_H = \{(\kappa, \theta, \gamma, \rho, \mu) \mid \kappa > 0,\, \theta > 0,\, \gamma > 0, \quad 0 < \theta < 1, \quad -1 \leq \rho \leq 1, \quad 2\kappa\theta > \gamma^2\}.$$

Define $U_s = X_{sT}$ and $V_s = Y_{sT}$. Then from Proposition 3.3.1 we have,

$$dU_s = T\mu U_s ds + \sqrt{T}\sqrt{V_s}U_s dZ_1(s) \tag{9.3}$$

$$dV_s = T\kappa(\theta - V_s)ds + \gamma\sqrt{TV_s}dB_1(s) \tag{9.4}$$

where Z_1 and B_1 are standard Brownian motions with correlation ρ.

9.1.1 Approximate log-likelihood function

Let $U_n = X_{nT}$ and $V_n = Y_{nT}$, $n = 0, 1, 2, \ldots, N$ be the given discrete observations of the asset price and square volatility respectively, where T is the time between consecutive observations. The Euler Maruyama discretization of the SDEs (9.3) and (9.4) leads to the following approximation,

$$U_{n+1} \approx U_n + T\mu U_n + \sqrt{T}\sqrt{V_n}\Delta Z_1(n), \quad n = 0, 1, 2, \ldots, N-1, \quad U_0 = X_0 = x_0 \tag{9.5}$$

$$V_{n+1} \approx V_n + T\kappa(\theta - V_n) + \gamma\sqrt{T}\sqrt{V_n}\Delta B_1(n), \quad n = 0, 1, 2, \ldots, N-1, \quad V_0 = Y_0 = y_0 > 0 \tag{9.6}$$

where

$$\Delta Z_1(n) = Z_1(n+1) - Z_1(n), \quad \Delta B_1(n) = B_1(n+1) - B_1(n), \quad E[\Delta Z_1(n)\Delta B_1(n)] = \rho T$$

and $y_0 > 0$ is a given fixed constant. We will obtain an estimate of H such that the discrete approximation (9.5)-(9.6) of the Heston model best fits the given data in a maximum likelihood sense.

Define,

$$Q(n) = \begin{pmatrix} Q_1(n) \\ Q_2(n) \end{pmatrix} = \begin{pmatrix} \Delta Z_1(n) \\ \gamma \Delta B_1(n) \end{pmatrix} \approx \begin{pmatrix} \dfrac{\Delta U(n) - T\mu U_n}{\sqrt{T}\sqrt{V_n}U_n} \\ \dfrac{\Delta V(n) - T\kappa(\theta - V(n))}{\sqrt{T}\sqrt{V_n}} \end{pmatrix},$$

Then for $n = 0, 1, 2, \ldots, N-1$, $Q(n)$ is a bi-variate normal random variable with mean M and variance-covariance matrix Σ given by,

$$M = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & \rho\gamma \\ \rho\gamma & \gamma^2 \end{pmatrix}.$$

The density function of $Q(i)$ is given by

$$(2\pi)^{-1}|\Sigma|^{-1/2} e^{-\frac{1}{2}Q(i)'\Sigma^{-1}Q(i)},$$

where $Q(i)'$ denotes the transpose of $Q(i)$ and $|\Sigma|$ denotes the determinant of the matrix Σ. The value of $|\Sigma|$ is $\gamma^2(1-\rho^2)$. From the Markov property of U_n and V_n, it follows that the approximate log-likelihood function $\tilde{L}_J$ with respect to the discrete dynamics (9.5)-(9.6) is given up to a constant by

$$\begin{aligned} \tilde{L}_J =& -N\ln(2\pi) - \frac{N}{2}\ln(\gamma^2(1-\rho^2)) - \frac{1}{2}\sum_{n=0}^{N-1}(Q(i)'\Sigma^{-1}Q(i)), \\ \frac{1}{N}\tilde{L}_J =& -\ln(2\pi) - \frac{1}{2}\ln(\gamma^2(1-\rho^2)) - \frac{1}{2N\gamma^2(1-\rho^2)}\sum_{n=0}^{N-1}(\gamma^2 Q_1^2(n) - 2\rho\gamma Q_1(n)z_2(n) + Q_2^2(n)). \end{aligned} \tag{9.7}$$

We equivalently minimize $L_J = -\tilde{L}_J$ to obtain the maximum likelihood estimators. The log-likelihood $\tilde{L}_J$ is a differentiable function of the parameters. If the parameter estimators are inside the open set $\mathcal{C}_H$ the gradient of the log-likelihood function must be zero at the

maximizer. This leads to the following first order conditions

$$\frac{\partial L_J}{\partial \mu} = \frac{1}{N}\sum_{n=0}^{N-1}((\gamma^2 Q_1(n) - \rho\gamma Q_2(n))\frac{\partial Q_1(n)}{\partial \mu}) = 0, \tag{9.8}$$

$$\frac{\partial L_J}{\partial \theta} = \frac{1}{N}\sum_{n=0}^{N-1}((Q_2(n) - \rho\gamma Q_1(n))\frac{\partial Q_2(n)}{\partial \theta}) = 0, \tag{9.9}$$

$$\frac{\partial L_J}{\partial \kappa} = \frac{1}{N}\sum_{n=0}^{N-1}((Q_2(n) - \rho\gamma Q_1(n))\frac{\partial Q_2(n)}{\partial \kappa}) = 0, \tag{9.10}$$

$$\frac{\partial L_J}{\partial \gamma} = \gamma^2(1-\rho^2) + \rho\gamma\frac{1}{N}\sum_{n=0}^{N-1} Q_1(n)Q_2(n) - \frac{1}{N}\sum_{n=0}^{N-1} Q_2^2(n) = 0, \tag{9.11}$$

$$\frac{\partial L_J}{\partial \rho} = \rho(1-\rho^2)\gamma^2 + \gamma(1+\rho^2)\frac{1}{N}\sum_{n=0}^{N-1} Q_1(n)Q_2(n) - \rho\sum_{n=0}^{N-1}(\gamma^2 Q_1^2 + Q_2^2) = 0. \tag{9.12}$$

9.1.2 Maximum likelihood estimators

The first-order conditions lead to non-linear algebraic equations in the 5 model parameters. Closed form solutions in terms of the observations only cannot be obtained for all parameters by solving simultaneously the first-order conditions. We will solve first for the estimator of μ using approximate log-likelihood based on (9.1) only. The estimator of μ based on (9.1) only was obtained in chapter 3. We recall $\hat{\mu}$ below,

$$\hat{\mu} = \frac{2}{NTd}\sum_{n=0}^{N-1}\frac{\Delta U_n}{U_n V_n}. \tag{9.13}$$

We solve the first order conditions (9.9)-(9.10) to obtain

$$\begin{aligned}\hat{\kappa} =& \frac{2\sqrt{T}\rho\gamma\zeta - \frac{2}{N}(V_N - V_0) + 2g}{T(f - \frac{4}{d})},\\ \hat{\theta} =& \frac{g}{\kappa T} + \frac{2}{d}, \quad \text{where } g = \frac{-b - 2\rho\gamma\tilde{\zeta}}{d},\end{aligned} \tag{9.14}$$

where

$$\zeta = \frac{1}{N}\sum_{n=0}^{N-1} Q_1(n)\sqrt{V_n}, \quad \tilde{\zeta} = \frac{1}{N}\sum_{n=0}^{N-1} \frac{Q_1(n)}{\sqrt{V_n}},$$

and we recall the statistics b, d and f from chapter 3,

$$b = -\frac{2}{N}\sum_{n=0}^{N-1} \frac{\Delta V_n}{V_n}, \quad d = \frac{2}{N}\sum_{n=0}^{N-1} \frac{1}{V_n}, \quad f = \frac{2}{N}\sum_{n=0}^{N-1} V_n.$$

We see that the estimators for both κ and θ depend on μ through $Q_1(n)$. We replace μ by $\hat{\mu}$ in $Q_1(n)$ to obtain $\hat{Q}_1(n)$ and use it in (9.14) to obtain $\hat{\kappa}$ and $\hat{\theta}$.

The first-order condition for ρ is cubic in ρ and depends on the parameters κ, θ and γ. After replacing κ and θ by $\hat{\kappa}$ and $\hat{\theta}$ respectively in the first-order condition for ρ we get a non-linear expression in ρ and γ. Similarly we get a non-linear expression in ρ and γ from the first order condition for γ. These two equations cannot be solved explicitly to obtain $\hat{\gamma}$ and $\hat{\rho}$ in terms of the observations only. We therefore using numerical optimization to obtain these two estimators. After obtaining $\hat{\gamma}$ and $\hat{\rho}$ numerically we use 9.14 to obtain the estimators for κ and θ. In our numerical implementation we use the inbuilt Matlab function *fminsearch* to minimize the log-likelihood as a function of ρ and γ. The *fminsearch* algorithm uses the Nelder-Mead simplex algorithm as described in [53].

9.2 Numerical comparison of the two estimation methods

We compare the joint parameter estimation method described in this chapter to the decoupled estimation described in chapter 3. We compare the two estimation methods for estimation accuracy and for the computation times. We fix the parameters of the Heston model corresponding to the 2006 S&P 500 data as the target parameter values that we are

trying to estimate. We recall below the value of the model parameters,

$$\kappa = 16.6, \quad \theta = .017, \quad \gamma = .28, \quad \rho = -.54. \tag{9.15}$$

Since the parameter μ is estimated using the same likelihood function in both cases we do not compare the results for μ.

To compare the accuracy between the two estimation methods we report the empirical bias and error in the parameter estimates from a sample of 1000 trajectories. The estimation error was defined in (5.3). We obtain an estimate of the error (5.3) by replacing the bias of the estimator by the empirical bias and standard deviation of the estimator by the empirical standard deviation computed over the 1000 trajectories. The empirical bias is computed as the difference of the sample mean of the estimators and the target value in (9.15). We give results for $N = 252,\ N = 504$ and $N = 1008$ with $T = 1/252$ in Table 9.1 and Table 9.2.

	Joint Estimation			Decoupled Estimation		
	N=252	N=504	N=1008	N=252	N=504	N=1008
$\hat{\kappa}$	.39	.62	.9	3.6	1.3	.55
$\hat{\theta}$	.0004	.0001	.0001	.00002	.00004	0
$\hat{\gamma}$	.01	.01	.01	.01	.01	. 01
$\hat{\rho}$	.0023	.0007	.0009	.0023	.0001	.0006

Table 9.1: Comparison of absolute bias between the estimators from joint estimation and decoupled estimation. Parameter values - $\kappa = 16.6,\ \theta = .017,$ $\gamma = .28,\ \rho = -.54,\ \mu = .102,\ T = .004.$

In Table 9.3 we report the CPU time in seconds to generate a sample of 1000 estimators of the five Heston model parameters from $N = 252, 504$ and $N = 1008$ observations from

	Joint Estimation			Decoupled Estimation		
	N=252	N=504	N=1008	N=252	N=504	N=1008
$\hat{\kappa}$	1.30	1.4	1.62	7.6	4.6	3.2
$\hat{\theta}$	.003	.002	.001	.002	.002	.001
$\hat{\gamma}$	.012	.012	.010	.015	.012	. 010
$\hat{\rho}$	.042	.023	.021	.058	.040	.030

Table 9.2: Comparison of standard deviation between the estimators from joint estimation and decoupled estimation. Parameter values - $\kappa = 16.6$, $\theta = .017$, $\gamma = .28$, $\rho = -.54$, $\mu = .102$, $T = .004$.

the two estimation methods. We conclude from these results that while the estimation from joint density is numerically tenuous there is no real advantage in terms of estimation accuracy.

	Joint Estimation	Decoupled Estimation
$N = 252$	22	.3
$N = 504$	83	.4
$N = 1008$	114	.4

Table 9.3: CPU-time (in seconds) for estimation of parameters from joint estimation and decoupled estimation

Bibliography

[1] Y. Achdou and O. Pironneau. *Computational methods for option pricing*. Society for Industrial Mathematics, 2005.

[2] Y. Aït-Sahalia, R. Kimmel, and F. Hall. Maximum likelihood estimation of stochastic volatility models. *NBER Working Paper*, 2004.

[3] Á. Alejandro-Quiñones, K. Bassler, M. Field, J. McCauley, M. Nicol, I. Timofeyev, A. Török, and G. Gunaratne. A theory of fluctuations in stock prices. *Physica A: Statistical Mechanics and its Applications*, 363(2):383–392, 2006.

[4] R. Azencott. Densité des diffusions en temps petit: développements asymptotiques. *Séminaire de Probabilités XVIII 1982/83*, pages 402–498, 1984.

[5] R. Azencott and D. Dacunha-Castelle. *Series of irregular observations: forecasting and model building*. Springer, 1986.

[6] R. Azencott and Y. Gadhyan. Accurate parameter estimation for coupled stochastic dynamics. In *Dynamical Systems and Differential Equations. Proceedings of the 7th AIMS international conference, Arlington, Texas, DCDS Supplement*, pages 44–53, 2009.

[7] R. Azencott, Y. Gadhyan, and R. Glowinski. Option price sensitivity to errors in stochastic dynamics modeling. In *Proceedings of the fourth SIAM conference on mathematics for industry*, pages 150–161, 2009.

[8] L. Bachelier. *Théorie de la spéculation*. Gabay, 1995.

[9] G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049, 1997.

[10] G. Bakshi, C. Cao, and Z. Chen. Do call prices and the underlying stock always move in the same direction? *Review of Financial Studies*, 13(3):549, 2000.

[11] O. Barndorff-Nielsen, E. Nicolato, and N. Shephard. Some recent developments in stochastic volatility modelling. *Quantitative Finance*, 2(1):11–23, 2002.

[12] D. Bates. Testing option pricing models. *NBER working paper*, 1995.

[13] T. Björk. *Arbitrage theory in continuous time*. Oxford Univ Press, 2009.

[14] F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 meetings of the business and economic statistics section, American Statistical Association*, volume 177, page 81, 1976.

[15] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 1973.

[16] C. Broto and E. Ruiz. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18(5):613–649, 2004.

[17] P. Carr and D. Madan. Option pricing and the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

[18] CBOE. http://www.cboe.com.

[19] M. Chernov and E. Ghysels. Estimation of stochastic volatility models for the purpose of option pricing. In *Computational Finance (Proceedings of the sixth international conference on computational finance), Leonard N. Stern School of Business*, 1999.

[20] M. Chernov and E. Ghysels. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56(3):407–458, 2000.

[21] K. Chung and R. Williams. *Introduction to stochastic integration*. Springer, 1990.

[22] L. Clewlow and X. Xu. The dynamics of stochastic volatility. *FORC preprint*, 94:53, 1993.

[23] J. Cox, J. Ingersoll Jr, and S. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 53(2):385–407, 1985.

[24] J. Cox, S. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

[25] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics An International Journal of Probability and Stochastic Processes*, 19(4):263–284, 1986.

[26] E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.

[27] G. Dohnal. On estimating the diffusion coefficient. *Journal of Applied Probability*, pages 105–114, 1987.

[28] J. Doob. *Stochastic processes*. New York, 1953.

[29] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[30] L. C. Evans. *Partial differential equations (Graduate Studies in Mathematics, V. 19) GSM/19*. American Mathematical Society, 1998.

[31] E. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36(4):420–429, 1963.

[32] W. Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182, 1951.

[33] J. Fouque, G. Papanicolaou, and K. Sircar. Mean-reverting stochastic volatility. *International Journal of Theoretical and Applied Finance*, 3(1):101–142, 2000.

[34] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Englewood Cliffs, NJ, 1964.

[35] C. Gardiner. *Handbook of stochastic methods*. Springer, Berlin, 1985.

[36] J. Gatheral and M. Lynch. Lecture 1: Stochastic volatility and local volatility. *Case Studies in Financial Modelling Course Notes, Courant Institute of Mathematical Sciences, New York University*, 2003.

[37] V. Genon-Catalot and J. Jacod. Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scandinavian Journal of Statistics*, 21(3):193–221, 1994.

[38] E. Ghysels, A. Harvey, and E. Renault. Stochastic volatily, in Handbook of Statistics 14, Statistical Methods in Finance. GS Maddala and CR Rao, 1996.

[39] I. Gihman and A. Skorohod. *Stochastic differential equations*. Springer-Verlag, 1972.

[40] R. Glowinski. *Handbook of numerical analysis*. North-Holland Amsterdam, 2003.

[41] R. Glowinski. *Numerical methods for nonlinear variational problems*. Springer-Verlag, 2008.

[42] J. Harrison and D. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381–408, 1979.

[43] D. Heath and M. Schweizer. Martingales versus PDEs in finance: an equivalence result with examples. *Journal of Applied Probability*, 37(4):947–957, 2000.

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

[45] D. Higham and X. Mao. Convergence of Monte Carlo simulations involving the mean-reverting square root process. *Journal of Computational Finance*, 8(3):35–62, 2005.

[46] L. Hörmander. *The Analysis of linear partial differential operators: Pseudo-differential operators*. Springer-Verlag, 1985.

[47] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300, 1987.

[48] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. North Holland, 1989.

[49] S. Ikonen and J. Toivanen. Operator splitting methods for American option pricing. *Applied Mathematics Letters*, 17(7):809–814, 2004.

[50] J. Jackwerth and M. Rubinstein. Recovering probability distributions from contemporaneous security prices. *Journal of Finance*, 51(5):1611–1631, 1996.

[51] M. Kendall and A. Stuart. *The advanced theory of statistics*. Macmillan, 1977.

[52] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin; New York, 1992.

[53] J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1999.

[54] C. Lamoureux and W. Lastrapes. Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Review of Financial Studies*, 6(2):293–326, 1993.

[55] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4), 1963.

[56] A. Melino and S. Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2):239–265, 1990.

[57] A. Melino and S. Turnbull. The pricing of foreign currency options. *Canadian Journal of Economics*, 24(2):251–281, 1991.

[58] R. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[59] P. Meyer. Un cours sur les intégrales stochastiques. *Séminaire de Probabilités 1967-1980*, pages 174–329, 2002.

[60] S. Mikhailov and U. Nögel. Hestons stochastic volatility model: implementation, calibration and some extensions. *Wilmott*, 2003(4):74–79, 2003.

[61] N. Moodley. The Heston model: A practical approach. *Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa*, 2005.

[62] C. Oosterlee. On multigrid for linear complementarity problems with application to American-style options. *Electronic Transactions on Numerical Analysis*, 15:165–185, 2003.

[63] S. M. Ross. *Introduction to probability models, Ninth Edition*. Academic Press, Inc., 2006.

[64] M. Rubinstein. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *Journal of Finance*, 40(2):455–480, 1985.

[65] M. Rubinstein. As simple as one, two, three. *Risk*, 8(1):44–47, 1995.

[66] P. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 6(2):13–32, 1965.

[67] L. O. Scott. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22(04):419–438, 1987.

[68] J. Singler. Differentiability with respect to parameters of weak solutions of linear parabolic equations. *Mathematical and Computer Modelling*, 47(3-4):422–430, 2008.

[69] K. Sircar and G. Papanicolaou. Stochastic volatility, smile & asymptotics. *Applied Mathematical Finance*, 6(2):107–145, 1999.

[70] E. Stein and J. Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4):727–752, 1991.

[71] R. Varga. Matrix Iterative Analysis. 2000.

[72] VIX White pages. http://www.cboe.com/micro/vix/vixwhite.pdf.

[73] J. Wiggins. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19(2):351–372, 1987.

[74] X. Xu and S. Taylor. The term structure of volatility implied by foreign exchange options. *Journal of Financial and Quantitative Analysis*, 29(01):57–74, 2009.

[75] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *Kyoto Journal of Mathematics*, 11(1):155–167, 1971.

[76] W. Yu, F. Zhang, and W. Xie. Differentiability of C0-semigroups with respect to parameters and its application. *Journal of Mathematical Analysis and Applications*, 279(1):78–96, 2003.