UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Understanding the complex dynamics of financial markets through microsimulation

Qiu, G.

Publication date
2011

Link to publication

Citation for published version (APA):
Qiu, G. (2011). *Understanding the complex dynamics of financial markets through microsimulation*.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:19 Sep 2021

Chapter 2

Empirical Observations and Limitations of Traditional Economic Theories

Financial economists and practitioners have recognized certain statistical regularity in financial time series, which is common across markets and time horizons, and known as 'stylized facts'. Importantly, they are contrary to the expectations of traditional economic and financial theories and challenge our understanding of their origins. This chapter presents a brief review of these empirical observations and their disagreement with mainstream financial theories.

2.1 Stylized facts observed in financial markets

Here we describe the stylized facts observed in two types of financial market, i.e. stock markets and option markets. In fact, these markets are among the most active financial markets with respect to trading volumes. In addition, most of the empirical studies of market dynamics in the literature were performed on the transaction data obtained from these two types of market.

2.1.1 Stock markets

The stylized facts of stock markets have been observed and discussed by many researchers. See, among many others, Cont (2001); Ding et al. (1993); Guillaume et al. (1997); Mandelbrot (1963); Mantegna and Stanley (2000); Pagan (1996); Voit (2003).

On long time scales (typically a week or longer), empirical distributions of financial return[1] generally fit to the Gaussian distribution. However, most financial returns over short timescales are described well by a non-Gaussian (heavy-tailed or fat-tailed) distribution.

A commonly used criterion for the normality of the distribution of a variable X is its kurtosis (κ), defined as

$$\kappa = \frac{E[(X-\mu)^4]}{\sigma^4}, \tag{2.1}$$

where μ and σ are respectively the mean and standard deviation of X. $\kappa = 3$ corresponds to a Gaussian distribution, whereas $\kappa > 3$ indicates a so-called leptokurtic distribution with a sharp peak and heavy tails.

The kurtosis of financial returns is far from that of a Gaussian distribution. For instance, our estimate for the kurtosis of S&P 500[2] daily returns over the period June 1950 to June 2005 is around 38. Figure 2.1(a) shows the distribution of these returns, together with a Gaussian probability density function (PDF)

[1] Generally, return is defined as $R_1^{t+1} = \ln P^{t+1} - \ln P^t$, where R_1^{t+1} is the return at time $t+1$, P^t is the price at time t, and so on. The basic relation is $R_2^{t+1} = (P^{t+1} - P^t)/P^t$. Sometimes, return is defined as price change, $R_3^{t+1} = P^{t+1} - P^t$. For high-frequency data, $|R_3^{t+1}| \ll P^t$. Hence, $R_1^{t+1} = \ln[1 + R_3^{t+1}/P^t] \simeq R_3^{t+1}/P^t = R_2^{t+1}$. Since R_3^t is a fast variable and P^t is a slow variable, $R_2^t \simeq CR_3^t$, where the time dependence of C is negligible. (See p. 35-39 of Mantegna and Stanley (2000).) So, $R_1^t \simeq R_2^t \simeq CR_3^t$. For high-frequency data, these three indicators are therefore alternatives to each other in analyzing the regularity in return distributions.

[2] An index is a sample list of stocks that is representative of a whole stock market. It is used by investors to track the performance of the stock market. Different methods are being used for calculating the price of an index. For example, the Dow Jones Industrial Average (DJIA), which contains 30 of the most influential companies in the U.S., is the price-based weighted average of the prices of the included stocks. The Standard and Poor's 500 Index (S&P 500), which includes 500 large publicly held companies that trade on major U.S. stock exchanges, weights companies by market capitalization (the overall value of a company's stock on the market).

and a Lorentz PDF for comparison. Clearly, daily returns of S&P 500 follow a non-Gaussian (fat-tailed) distribution, implying a greater frequency of extreme events than would be expected if they followed a normal distribution. However, the variance of the distribution is finite, whereas that of a Lorentz distribution (or a stable Lévy distribution in general) is infinite.

In statistics, the autocorrelation function (ACF) of a time series describes the correlation between the values of the series at two different points in time. The autocorrelation coefficient of a process X_t as a function of time lag τ is defined as

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2}. \tag{2.2}$$

Plotted against τ, $R(\tau)$ can be used for detecting the non-randomness of X_t.

The ACF of daily returns of financial assets quickly converges to the noise range, whereas the corresponding ACF of volatility[1] decays slowly (see Figure 2.1(b)). The long-term autocorrelation of volatility is the reflection of the phenomenon termed 'volatility clustering' — high (positive or negative) returns tend to group together. Figure 2.1(c) shows the time series of return over the period. In this figure, the effect of volatility clustering is clearly illustrated.

Based on a large amount of available transaction data, Mantegna and Stanley (1995) explored whether the scaling phenomena occur in financial markets. Specifically, they showed that the scaling of the probability distribution of S&P500 can be described by a non-Gaussian process. In addition, the scaling behavior can be observed for time intervals spanning three orders of magnitude (from 1,000 min to 1 min) and the scaling exponent is remarkably constant over the six-year period (1984-1989) of the data.

Remarkably, non-Gaussian (fat-tailed) distribution, volatility clustering, and the scaling law have also been found in the time series recorded in some other types of financial market, e.g., foreign exchange markets (Müller et al. (1990)).

[1] In the finance literature, volatility refers to the spread of asset returns measured as the standard deviation of a sample of returns over a period of time, i.e., $\sigma = \sqrt{(1/T)\sum_{t=1}^{T}(R^t - \overline{R})^2}$, where T is the length of the period, R^t is the return at time t, and $\overline{R}$ is the average return over the period. Substituting $T = 1$ and $\overline{R} = 0$ into this equation, we obtain the absolute value of the return over a period of one time unit, $|r|$, which is the most commonly used proxy for volatility in practice. The other commonly used proxy is r^2. We adopt $|r|$ as volatility.

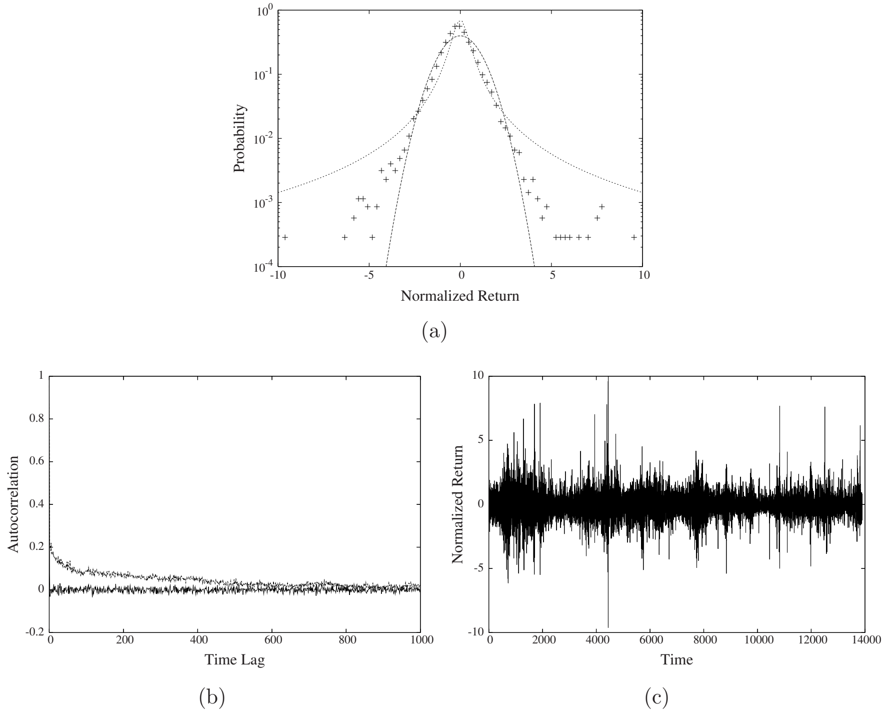


Figure 2.1: (a) Distribution of the daily returns of S&P 500 over the period from June 1950 to June 2005 (the points), compared with a Gaussian PDF (the curve that decays faster) and a Lorentz PDF. A logarithmic scale is used for the vertical axis. (b) Autocorrelation function of the daily returns (the lower line) and the corresponding ACF of volatility. (c) Time series of the daily returns.

In literature, there is not yet a common agreement on the origins of the stylized facts (Cont (2005)). On the other hand, the analytical models developed for describing these phenomena, some of which are described in Section 2.2.1, do not provide explicit economic explanations for the underlying mechanisms.

2.1.2 Options markets

The volatility smile phenomenon observed in options markets is a long-standing problem in financial economics and has been discussed by many researchers (Bakshi et al. (1997); Black (1975); Ciliberti et al. (2009); Cont and da Fonseca (2002); Ederington and Guan (2002); Fengler et al. (2003); Geman (2005); Johnson et al. (2003); Macbeth and Merville (1979); Rebonato (2004); Rubinstein (1985); Tompkins (2001)).

This phenomenon is intrinsically related to the widely used Black-Scholes (BS) model for option pricing (Black and Scholes (1973), Merton (1973)). The BS formula for pricing a European option[1] on a non-dividend-paying asset is

$$V_{BS}^{\phi,t}(S^t, K, r, \tau, \sigma) = \phi[S^t N(\phi d_1) - Ke^{-r\tau} N(\phi d_2)], \tag{2.3}$$

where S^t is the price of the underlying asset at time t, K the strike price of the option, r the risk-free interest rate, $\tau = T - t$ the time to maturity where T the expiration time of the option, σ the volatility of the asset, and $\phi = 1(-1)$ for a call (put) option. In Equation (2.3), d_1 and d_2 are defined as

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

$$d_2 = d_1 - \sigma\sqrt{\tau}, \tag{2.5}$$

and $N(x)$ is the standard normal cumulative distribution function.

All parameters in the BS model other than the volatility are observables and, according to the model, the theoretical value of an option is a monotonic increasing function of the volatility. A unique volatility is therefore implied by the market price of an option ($V^{\phi,t}$), the so-called *implied volatility*:

$$\sigma_{imp}^{\phi,t} : \quad V_{BS}^{\phi,t}(S^t, K, r, \tau, \sigma_{imp}^{\phi,t}) = V^{\phi,t}. \tag{2.6}$$

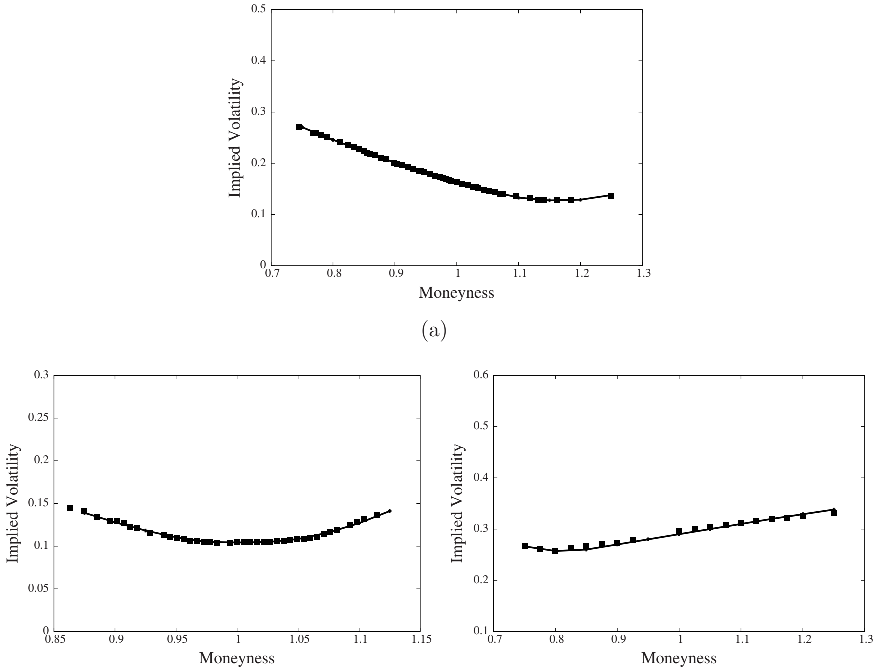
[1] A European option may be exercised only at the expiry date of the option; in contrast, an American option may be exercised at any time before the expiry date (Hull (2003)).

According to the BS model, implied volatility (IV) is independent of strike for a fixed time to maturity. Hence, plots of IV against strike should be flat. In reality, however, it is well known that IVs exhibit a remarkable curvature, which is commonly referred to as a *volatility smile*.

In particular, equity index options tend to have a downward sloping IV curve, i.e. a volatility skew. This skew has become much more pronounced after the stock market crash of October 1987. Foreign currency options, however, typically show a symmetric smile, especially if the currencies are of equal strength. Contrary to equity index options, some commodity options often show an upward sloping skew. Figure 2.2 depicts the IVs and the fitted curves of options derived from three underlying assets of different types: Equity index, currency, and commodity. The downward IV skew shown in Figure 2.2(a) is obtained from the European-style options on the S&P500 index with time to maturity 118 days, traded on April 7, 2004 in Chicago Mercantile Exchange. The symmetric IV smile shown in Figure 2.2(b) is from the British Pound options with time to maturity 43 days, traded on April 7, 2004 in Chicago Mercantile Exchange. The upward IV skew shown in Figure 2.2(c) is from the soybeans options with time to maturity 137 days, traded on January 15, 2004 in Chicago Board of Trade.

Another important fact is that the volatility smile changes over time. It has been revealed that three principal components can sufficiently account for the observed deformation of the smile: The first component reflects the shift in its overall level; the second one conveys the change of its slope; and the third component accounts for the adjustment in its convexity (Cont and da Fonseca (2002); Fengler et al. (2003)). Figure 2.3 shows the principal components obtained from some empirical time series of implied volatility and the proportions of the variances explained by these principal components. The results shown in Figures. 2.3(a) and 2.3(b) are based on those reported in Fengler et al. (2003), which were obtained from the daily IV time series of European style options on DAX index for the entire year 1999. The results shown in Figures. 2.3(c) and 2.3(d) are based on those reported in Rexhepi (2008), generated from the daily IV time series of options on S&P500 index over the period from December 1999 to October 2002.

There are other stylized facts or observed characteristics with respect to the IV dynamics. For detailed discussions, see Cont and da Fonseca (2002), Rebonato



(b) (c)

Figure 2.2: Empirical implied volatilities plotted against moneyness (K/S^t), for different types of underlying asset. (a) Implied volatilities (the points) and the fitted curve of the European-style options on the S&P500 index with time to maturity 118 days, traded on April 7, 2004 in Chicago Mercantile Exchange. (b) Implied volatilities (the points) and the fitted curve of the British Pound options with time to maturity 43 days, traded on April 7, 2004 in Chicago Mercantile Exchange. (c) Implied volatilities (the points) and the fitted curve of the soybeans options with time to maturity 137 days, traded on January 15, 2004 in Chicago Board of Trade. Notice that the three data sets are different in the range of moneyness and that of implied volatility.

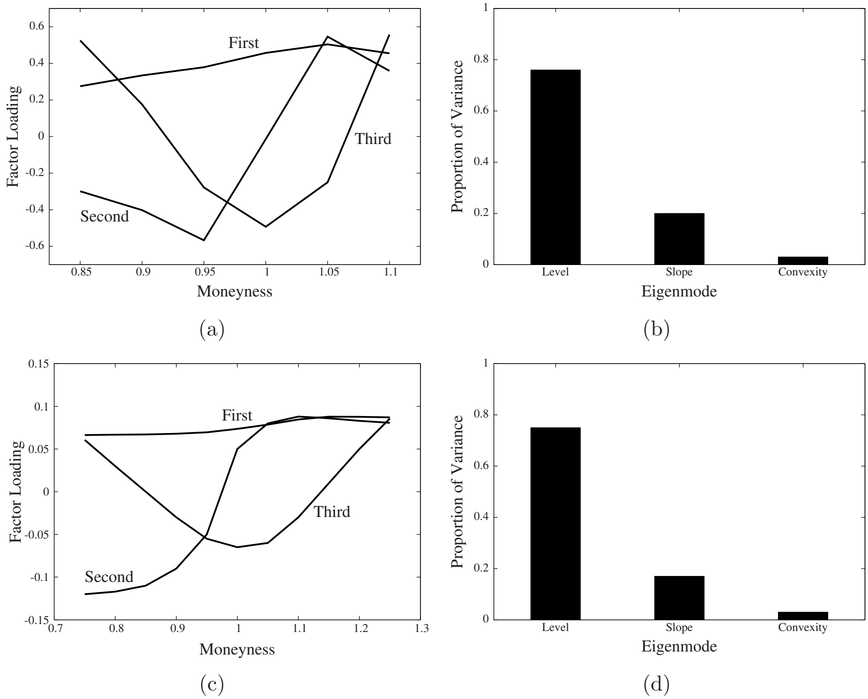


Figure 2.3: Empirical principal components and their proportions of variance. (a) Principal components calculated using the daily IV time series of European style options on DAX index for the entire year 1999, reported by Fengler et al. (2003); (b) Corresponding proportions of variance explained by the principal components. (c) Principal components obtained from the daily IV time series of options on S&P500 index over the period from December 1999 to October 2002, calculated by Rexhepi (2008); (d) Corresponding proportions of variance explained by the principal components. Notice that (a) and (c) are different in the range of moneyness and that of factor loading.

(2004), and Ederington and Guan (2002), among many others.

The existence of a volatility smile conflicts with the BS framework. To account for the associated deviations of option prices from the BS formula, models based on processes for underlying assets other than geometric Brownian motion[1] such as stochastic volatility and jump diffusion processes, have been proposed (Hull, 2003; Rebonato, 2004). Some models of this type are described in Section 2.2.2. They can include the smile effect on the valuation of options (Cont and Tankov, 2004; Gatheral, 2006). However, although these models may agree on option prices today, they can differ in the future prices (Bakshi et al., 1997). This disagreement on the so-called implied smile dynamics results in different prices of exotic options. In addition, it is important to realize that although traders may agree on the best model to use for a certain product, they can still disagree about the inputs of the model such as the volatility parameter mentioned above. This will lead to different prices for the same option contract where clients will most likely select the best offer, indicating that the market prices of options are ultimately determined by supply and demand. Finally, an important fact to note is that the volatility skew in the equity option market has only been observed after the stock market crash of October 1987 (Hull, 2003), while the non-Gaussian distribution of price changes[2] has been empirically identified in financial time series accumulated since at least the beginning of the 20th century (Mandelbrot, 1963). From these perspectives, the adoption of generalized stochastic processes in option pricing may not explain the main cause of the smile. Therefore, the origin of this phenomenon remains unclear.

During the last few decades, a series of formal studies have been undertaken to explain market anomalies based on the behavior of market participants and some of the work involves the volatility smile. Recent years have also witnessed active investigations of market dynamics through microsimulation, and a few MS-based studies address the smile phenomenon. Some of these models are introduced in

[1] In the BS model, a geometric Brownian motion is assumed for the price dynamics of the underlying asset.

[2] The geometric Brownian motion implies a Gaussian distribution of price changes, while stochastic volatility and jump diffusion processes indicate a non-Gaussian distribution.

Section 3. While these studies have clearly demonstrated the potential of behavioral and agent-based approaches for studying the origin of the smile phenomenon, they have not reproduced or realistically explained IV curves of options on different types of underlying asset, such as the upward sloping skew observed in commodity options markets. In addition, they have not confirmed the dynamical properties of the smile.

2.2 Empirically consistent models

In the literature, some analytical models have been developed in order to capture the complex dynamics observed in asset prices. In this section, we describe those that are most widely applied.

2.2.1 Models of price dynamics

The ubiquity of fat-tail distributions and volatility clustering has stimulated a great deal of theoretical work for developing models more consistent with empirical time series. Among the many models of this type, the ARCH model that was introduced by Engle (1982) and the GARCH model proposed by Bollerslev (1986) are the most widely used.

ARCH stands for 'autoregressive conditional heteroscedasticity'. To be specific, an ARCH(q) process is defined as

$$\sigma^2(t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2(t-i), \tag{2.7}$$

with

$$\varepsilon(t) \sim \mathcal{N}(0, \sigma^2(t)). \tag{2.8}$$

Here $\mathcal{N}(0, \sigma^2)$ represents a normal distribution with mean 0 and standard deviation σ, $\alpha_i (i = 0, 1, ..., q)$ are positive parameters, and the random variable $\varepsilon(t)$ is drawn from a normal distribution with mean 0 and time-dependent standard deviation $\sigma(t)$. Here $\sigma^2(t)$ is determined by the last q realizations of $\varepsilon(t)$.

GARCH represents ‘generalized autoregressive conditional heteroscedasticity’. A GARCH(p, q) process is defined as

$$\sigma^2(t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2(t-i) + \sum_{i=1}^{p} \beta_i \sigma^2(t-i), \tag{2.9}$$

where $\sigma^2(t)$ is additionally dependent on its last p values.

By varying the value of q, we can control the memory effect. By including the memory of $\sigma^2(t)$ itself, the GARCH model can overcome the difficulties of the ARCH model in the optimal determination of the $q+1$ parameters (Mantegna and Stanley (2000)). Importantly, although the process of $\varepsilon(t)$ is chosen to be Gaussian, the asymptotic PDF presents a degree of leptokurtosis. Therefore, ARCH and GARCH models are simple models able to produce empirically consistent time series of asset price. Bera and Higgins (1993) remarked that a major contribution of the ARCH/GARCH literature is the finding that the changes in volatility may be predictable and result from a specific type of nonlinear dependence rather than changes in exogenous variables.

Through parameter estimations and variations on the basic model, ARCH and GARCH models can be effective for forecasting volatility. Therefore, they have widely and successfully applied in finance, of which the volatility is a central issue because financial decisions are generally based upon the tradeoff between risk and return.

However, these models are not compatible with all of the empirical properties of price fluctuations (Farmer (1999)). Conventional ARCH-type models are incompatible with the scaling properties of price fluctuations: They may fit at a given timescale but do not work well for explaining the volatility at a different timescale (Farmer (1999) and Mantegna and Stanley (2000)). In addition, conventional ARCH models do not have asymptotic power-law decay in the volatility autocorrelation function presented in empirical financial time series (Farmer (1999), Mantegna and Stanley (2000), Cont (2005)).

2.2.2 Option pricing models

The volatility smile phenomenon has attracted considerable attention in financial economics. In order to more realistically price options, many new alternatives to the Black-Scholes model have been proposed. In these models, some of the restrictive assumptions of the BS framework have been relaxed. For example, they allow volatilities, interest rates, or price jumps to be stochastic.

However, Bakshi et al. (1997) empirically examined several alternative models and concluded that, judged on consistency of implied parameters, all these models are misspecified. For example, the required level of volatility variation is implausibly too high. Rebonato (2004) reported that financially plausible models cannot produce good fit to the market prices of options, while models that do generate the smile are not financially convincing. In addition, various models are often combined in order to achieve a better fit to market data. In this case, models are no longer parsimonious, and it might be difficult to judge whether a good fit indicates a well-specified modeling approach, or is simply the result of the flexibility brought by the larger number of parameters.

Here we describe and examine some specific alternative efforts that focus on refining the volatility process. There are two broad approaches of this type. In *local volatility* models (Rebonato (2004)), the volatility process is represented by a deterministic function of the asset price and time,

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dz, \tag{2.10}$$

in which the terms $\mu(S_t, t)$ and $\sigma(S_t, t)$ are respectively the drift and the volatility of the stochastic process, and dz is a Wiener process or Brownian motion.

In *stochastic volatility* models, the volatility dynamics is described as a second stochastic process V_t in addition to the price process S_t,

$$dS_t = \mu_S(S_t, \sqrt{V_t}, t)dt + \sigma_S(S_t, \sqrt{V_t}, t)dz_1, \tag{2.11}$$

$$dV_t = \mu_V(S_t, V_t, t)dt + \sigma_V(S_t, V_t, t)dz_2, \tag{2.12}$$

$$E[dz_1 dz_2] = \rho dt \tag{2.13}$$

where dz_1 and dz_2 are two Wiener processes with constant correlation factor ρ.

We describe two typical local volatility models, i.e. the displaced diffusion model that was developed by Rubinstein (1983) and the constant elasticity of variance (CEV) model proposed by Cox and Ross (1976), and one typical stochastic volatility model, i.e. the SABR (Stochastic Alpha, Beta, Rho) model introduced by Hagan et al. (2002). In addition, we show the experimental results obtained by using these models, with regard to their in-sample and out-of-sample performances. The former shows how well they fit the empirical IV curve of each day; the latter indicates how well the models can fit the empirical curves of other days, by using the parameters obtained from the empirical data today.

Displaced diffusion model

Fundamental to the displaced diffusion model is the displaced-diffusion process requiring that the quantity $S_t + a$, rather than S_t, should follow a geometric Brownian motion:

$$\frac{d(S_t + a)}{S_t + a} = \mu_a dt + \sigma_a dz_t, \tag{2.14}$$

where μ_a and σ_a are respectively the percentage drift and volatility of $S_t + a$, in which a is a positive constant termed the displacement coefficient.

Due to the properties of geometric Brownian motion, only $S_t + a$, rather than S_t, is guaranteed to be positive. Since $a > 0$, we have $-a < S_t < +\infty$, indicating an obvious drawback of this model: S_t can be negative, in odd with the definition that S_t is the price of the underlying which should be strictly positive.

Leaving out the drift term, we start from

$$dS_t = \sigma_{abs} dz_t + \sigma_{log} S_t dz_t, \tag{2.15}$$

where σ_{abs} is the absolute responsiveness to the Brownian motion and σ_{log} is the responsiveness to the Brownian motion proportional to S_t. Equation (2.15) can be written as

$$dS_t = \sigma_{log}(S_t + \frac{\sigma_{abs}}{\sigma_{log}}) dz_t. \tag{2.16}$$

Since $d(S_t + a) = dS_t$, Equation (2.16) can be expressed as

$$\frac{d(S_t + \frac{\sigma_{abs}}{\sigma_{log}})}{S_t + \frac{\sigma_{abs}}{\sigma_{log}}} = \sigma_{log} dz_t. \tag{2.17}$$

Equation (2.14) is obtained by equating $\frac{\sigma_{abs}}{\sigma_{log}}$ with a and σ_{log} with σ_a. It can be seen that as a goes to zero, the process will behave similarly to a log-normal process and as a goes to infinity the process becomes a normal distribution. Option prices are given by the BS formula with the exception that the stock price is $S_t + a$ and the strike is $K + a$,

$$C_{BS}(S_t + a, t) = (S_t + a)N(d_1) - (K + a)e^{r\tau}N(d_2) \tag{2.18}$$

$$d_1 = \frac{\ln \frac{S_t+a}{K+a} + (r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \tag{2.19}$$

$$d_2 = d_1 + \sigma\sqrt{\tau} \tag{2.20}$$

where all the other parameters are identical to those in Equation (2.3).

As shown by Rebonato (2004) and Rexhepi (2008), the IV curves produced by the displaced diffusion model are downward sloping. To some extent, they agree with the empirical IV curves of options on equities. However, in order to fit the model to the empirical skews, the value of the displacement coefficient a is typically larger than the value of the underlying asset S_t, unrealistically indicating that S_t follows a Gaussian-like distribution with a large probability of taking a negative value. In addition, both the in-sample and out-of-sample performance of this model are poor (Rexhepi (2008)).

Constant elasticity of variance model

Inspired by the empirical fact that prices of underlying assets and corresponding implied volatilities display a strong inverse relationship, the so-called *leverage effect*, the constant elasticity of variance (CEV) model assumes that the underlying follows the following process,

$$dS_t = rS_t dt + \sigma S_t^{\alpha} dz_t, \tag{2.21}$$

where r is the risk-free interest rate, σ is a volatility parameter, α is a positive constant, and dz_t is a Wiener process.

The volatility of the underlying is therefore $\sigma S_t^{\alpha-1}$. When $\alpha = 1$, the process reduces to a geometric Brownian motion. When $\alpha < 1$, the volatility increases as the price of the underlying decrease. This creates a probability distribution similar to that observed in equity markets which is characterized by a heavy left tail and a less heavy right tail: As the price of the underling decreases, the volatility increases, making even lower prices more likely; when the price of the underling increases, the volatility decreases, making higher prices less likely (Hull (2003)). This produces the leverage effect observed in equity options.

The CEV option pricing formula can be expressed in terms of the non-central chi-square distribution. For $0 < \alpha < 1$, prices of call and put options are

$$c = S_0 e^{-qT}[1 - \chi^2(a, b+2, c)] - Ke^{-rT}\chi^2(c, b, a), \tag{2.22}$$

$$p = Ke^{-rT}[1 - \chi^2(c, b, a)] - S_0 e^{-qT}\chi^2(a, b+2, c), \tag{2.23}$$

where

$$a = \frac{K^{2(1-\alpha)}}{(1-\alpha)^2\sigma^2 T}, \quad b = \frac{1}{1-\alpha}, \quad c = \frac{(Se^{(r-q)T})^{2(1-\alpha)}}{(1-\alpha)^2\sigma^2 T}. \tag{2.24}$$

When $\alpha > 1$,

$$c = S_0 e^{-qT}[1 - \chi^2(c, -b, a)] - Ke^{-rT}\chi^2(a, 2-b, c), \tag{2.25}$$

$$p = Ke^{-rT}[1 - \chi^2(a, 2-b, c)] - S_0 e^{-qT}\chi^2(c, -b, a). \tag{2.26}$$

Here $\chi^2(z, v, k)$ is the cumulative probability that a variable with a non-central χ^2 distribution with non-centrality parameter v and k degrees of freedom is less than z.

As shown in Rexhepi (2008), the CEV model can generate a IV skew when the value of α is small. The skew flattens out as the α value increases. This model performs better than the displaced diffusion model in in-sample fitting, although still poorly because the skewness obtained is much smaller that the empirical one. In addition, the model performs poorly in out-of-sample fitting.

SABR model

The SABR model describes a forward price of the underlying F_t, of which the volatility is described by a parameter σ_t. The time evolution of F_t and σ_t is defined as the following system of stochastic differential equations,

$$dF = \omega_t F_t^{\beta} dz_1, \tag{2.27}$$

$$d\omega_t = \alpha\omega_t dz_2, \tag{2.28}$$

$$E[dz_1 dz_2] = \rho dt, \tag{2.29}$$

with

$$F_0 = f, \tag{2.30}$$

$$\omega_0 = \alpha, \tag{2.31}$$

where ω_t is the volatility process and ρ is the correlation coefficient of the two Wiener processes.

According to the SABR model, the price of a European option is given by the Black-Scholes formula

$$\begin{aligned} C_{BS}(S^t, t) &= S^t N(\phi d_1) - K e^{-r\tau} N(\phi d_2), \\ d_1 &= \frac{ln(S^t/K) + (r + \sigma_B^2/2)\tau}{\sigma_B \sqrt{\tau}}, \\ d_2 &= d_1 - \sigma_B \sqrt{\tau}, \end{aligned} \tag{2.32}$$

where S^t, K, r, τ, and $N(x)$ are identical to the corresponding terms in Equation (2.3). The implied volatility σ_B can be approximated as,

$$\begin{aligned} \sigma_B(f, K) = \frac{\alpha}{f^{1-\beta}} \{1 - \frac{1}{2}(1 - \beta - \rho\lambda) \log(\frac{K}{f}) + \frac{1}{12}[(1-\beta)^2 \\ +(2 - 3\rho^2)\lambda^2] \log^2 \frac{K}{f} + \cdots \}, \end{aligned} \tag{2.33}$$

provided that the strike price K is not too far from the current forward (Hagan et al. (2002)). The parameter

$$\lambda = (\frac{\upsilon}{\alpha}) f^{1-\beta} \tag{2.34}$$

is the ratio that measures the volatility of volatility at the current forward.

The SABR model has four parameters: α, β, υ, and ρ. α mainly controls the overall level of the IV curve, ρ and β determines the skewness of the curve, and υ controls the convexity.

As shown in Rexhepi (2008), the SABR model can produce strong IV skews and can fit in-sample IV curves better than the previously described local volatility models. The main reason for the better fit lies in the fact that the SABR have more parameters, enabling it to more flexibly fit complicated curves. With respect to out-of-sample fitting, the SABR model performs very well in fitting the empirical data of the days not so far from the sample day, but poorly in fitting the empirical IV curves of the days longer than a few days apart.

2.3 Problems of standard financial theories

Neoclassical economics stands for a general approach to economics that relates supply and demand to representative market participants with rational preferences and expectations. It defines itself as primarily the study of the allocation of resources and presumes that the market mechanism will lead to a general equilibrium. This approach was developed in the late-nineteenth century and represented the dominant tradition of economic theory throughout most of the twentieth century. Neoclassical economics has received widespread criticisms that mainly concern its adoption of many unrealistic assumptions. (See Farmer (1999), Mantegna and Stanley (2000), Tesfatsion (2002), and Hommes (2006)).

Inheriting the characteristics of the earlier neoclassical economics, the standard (mainstream) finance, which was shaped in the second half of the twentieth century, made great achievements evidenced by the births of the capital asset pricing model (CAPM) (Sharpe (1964)), efficient markets hypothesis (EMH) (Fama (1970)), and the Black-Scholes option pricing model (Black and Scholes (1973); Merton (1973)), among many others.

CAPM is an economic model for determining the required return rate of an asset, based on the idea that the investor demands the time value of the investment and a risk premium. The former is represented by the risk-free rate, while the latter is the product of a risk measure, i.e. the sensitivity of the returns of the asset to market returns, and a market premium, i.e. the difference between the expected market return and the risk-free rate. If the expected return rate is not equal or greater than the required rate, the investor will not perform the investment.

EMH states that at any given time, security prices already reflect all known information, whereas future price movements are determined entirely by new information and cannot be predicted by using past price movements. Therefore, prices must follow a Markov process. Because price movements do not follow any patterns or trends, it is impossible to consistently outperform the market, except through luck or taking riskier investments.

The BS model, as described in Section 2.1.2, is used for pricing European put and call options. The fundamental idea behind this model is to replicate an option using positions in the underlying stock and bonds. The fair value of the option is equal to the cost of setting up such a replicating portfolio. Two important assumptions of this model are (1) markets are efficient and arbitrage-free[1], and (2) the returns on the underlying stock are normally distributed. The first assumption implies that people cannot consistently predict the direction of each stock and share prices following a Markov process. Loosely speaking arbitrage-free means that there are no trade strategies that will result in a sure profit with a zero investment today.

Frequently, implications and expectations of standard financial theories do not agree with empirical observations. For instance, CAPM and EMH expect steady price movements corresponding to a Gaussian return distribution with time-independent variance (Sharpe (1964)), while empirical returns follow a non-Gaussian, leptokurtic distribution and their absolute values are with temporal correlations, as described in Section 2.1.1. Yet another example is that the BS model expects identical implied volatilities for all options on the same underlying, whereas real implied volatilities are different across strike and change over time, as depicted in Section 2.1.2.

Two widely accepted explanations, among numerous plausible ones, for the deviations of standard theories from empirical observations are discussed below.

2.3.1 Adoption of unrealistic assumptions

The mainstream financial economics relies on a top-down construction based on a number of unrealistic assumptions mainly for the sake of analytical tractability (Farmer (1999), Mantegna and Stanley (2000), Tesfatsion (2002), Hommes (2006)).

Standard theories typically assume that the aggregate effect of the participants in a market can be replaced with that of a so-called *representative* agent who maximizes an expected utility function[2]. A representative agent is therefore an

[1]The no arbitrage condition states that in a financial market it should not be possible to make a profit with zero net investment and without bearing any risk.

[2]Briefly, utility is a synonym for individual welfare, i.e., satisfaction from consumption of goods and services; a utility function is a representation expressing the relationship between the utility and the consumption (Black et al. (2002))

idealized trader whose behavior represents the aggregate action of all agents. Through this aggregation, economists avoid the difficulties of modeling a group of traders and wish to derive the behavior of a whole market.

They also assume that traders are rational, i.e. they have clear preferences, make identical unbiased forecasts about the future, and perform actions to maximize their chances of success. In addition, traders are considered as efficient machines who have unlimited computing abilities enabling them to process any inflow of new information and make optimal choices instantaneously.

There are some other assumptions commonly made in standard financial theories. For example, the returns of traded assets follow a normal distribution, traders have identical beliefs, all relevant information is known to all parties involved, and markets are frictionless, etc. (Farmer (1999), Mantegna and Stanley (2000), Tesfatsion (2002), Hommes (2006)).

All these assumptions have received severe criticisms. For example, under the assumption that traders are rational and with homogeneous beliefs, all investors will hold the same market portfolio and there will be no trade, contradicting the large trading volumes recorded frequently in real markets. Another important point is that, it is very difficult to obtain analytical results if even one of the assumptions of a model is relaxed, and the more is a model adjusted to represent the real-life situation, the less is it analytically tractable. In addition, it is hard to figure out analytically the effect of the relaxation of each of the unrealistic assumptions of a model on its equilibrium results (Levy et al. (2000)).

Here we again take the BS model as an example to indicate that some of its assumptions are indeed unrealistic and how they affect real world financial operations, in particular, risk management. A detailed explanation of this point can be found in Johnson et al. (2003). The BS model is widely used for creating delta-neutral portfolios, i.e., portfolios that are insensitive to changes in the value of the underlying. For example, option writers are exposed to large potential losses, so they usually hedge their position by buying a certain quantity of the underlying asset. By keeping this portfolio neutral to changes in the price of the underlying, an option writer can hedge away the risk associated with the short position in the option. The quantity of underlying is determined by *delta*, which is represented as the partial derivative of the option's fair value with respect to the price of the underlying and usually calculated using the BS formula. Importantly, delta is a function of the price of the underlying and therefore changes over time. Consequently, the option writer's position remains delta neutral only for a short period of time and the hedge has to be adjusted periodically.

According to the BS model, it is therefore possible that, through this dynamic-hedging scheme, the variation of the option writer's wealth always remains zero. Some of the main assumptions adopted by the BS model that guarantee zero risk are (1) there are no transaction cost, (2) the trading is continuous, and (3) the price of the underlying follows a Geometric Brownian motion. In reality, however, all these assumption are questionable. Cost is always present and it gives rise to a barrier to high-frequency trading: The greater the frequency of re-hedging, the greater the cost. In addition, as discussed in Section 2.1, returns of the underlying do not follow a Gaussian distribution.

Bouchaud and Potters (2000) and Johnson et al. (2003) examined the variation of the option writer's wealth when the price of the underlying changes over time, by studying the basic components of the wealth. They focused on the two basic components, i.e., the payoff, which the option writer must give to the option holder at the expiry of the option, and the hedging profit, which comes from the profit or loss realized on the purchased quantity of the underlying asset. In order to investigate the effects of discrete hedging on the risk of writing an option, they simulated repeatedly the process of writing and hedging an option, under different schemes of hedging, different underlying asset movements, and different option types. At each re-hedging time, the hedge quantity of the underlying is calculated using the BS delta-hedging recipe. In the case that the price of the underlying follow random walk, the simulation results showed that as the frequency of re-hedging increases, the spread in the variation of wealth decreased, meaning less risk for the option writer. Specifically, as the trading time reduces to zero, the spread in the distribution of wealth variation also reduces to zero, recovering the BS result. However, by using a slightly more realistic model for the price movement of the underlying, e.g., a process with stochastic volatility, the simulation results showed a marked increase of risk for all trading times. Importantly, reducing the trading time to zero no longer gets the zero-risk result.

Johnson et al. (2003) further addressed the crucial issue of managing portfolio in the presence of non-zero transaction costs. The more frequently the option is hedged, the more risk can be eliminated, however, the more cost is paid. Hence, the option price can only be minimized by balancing the reduction of risk with the increase in transaction cost, but not by neglecting both of them as carried out in the deriving of the the BS model.

2.3.2 Evasion of endogenous mechanisms

By adopting a representative agent whose behavior represents the actions of all the traders in a market, standard economists have implicitly adopted the method of reductionism, through which the dynamics of a system is described simply as the sum of the dynamics of its components (Tesfatsion (2002), Hommes (2006)).

However, the integrated behavior of a complex system generally cannot be deduced by simply summing the behavior of the components. In fact, the most complex behavior of a system usually arises from the interactions among its components, not from the complexity of extraneous factors or that of the components themselves (Mantegna and Stanley (2000), Voit (2003)).

Financial markets are characterized by the completely lack of linearity. If we break down a market into individual traders who do not interact, the market ceases to exist. In addition, traders of distinct types are very different in trading interest and trading behavior, the aggregate effect of their actions can hardly be translated to that of any sensible types of representative agent. In studying the complex dynamics of financial markets, the approach of reductionism can hardly work.

In fact, some mainstream financial economists have realized the importance of the interactions between market participants and the feedback between micro and macro market structures. However, they have avoided modeling them explicitly. A reasonable explanation for this evasion is that, for a long period of time, economists lacked the means to handle the modeling of trading behavior in real markets (Tesfatsion (2002)). Economists have therefore turned to model a financial market based on its macro-level manifestations. However, economists in this track have realized that the same data might lead to wildly divergent models performing equally well and, after all, the conclusions based on such models are not robust (Sterman (2000)).

As a typical and most successful neoclassical model, the BS model is a *principle theory*, in that the various principles going into their model were given the status of postulates, and no underlying mechanisms for the phenomena were elucidated. By contrast, a *constructive theory* is derived from the data and are physically well-founded by providing basic mechanisms for the phenomena. Rickles (2008) explained the distinction between these two types of approach by citing thermodynamics as well as neoclassical economics (on the 'principle' side), and statistical mechanics as well as econophysics (the 'constructive' side).

In deriving principle theories, some general principles that were assumed to be universally valid are taken and conclusions are drawn about the nature of the object under study based on these universal assumptions. In physics, principle theories, such as Einstein's theory of relativity, were not hypothesis built on data reached through experimentation, but were universal principles intended to impact all of physics. However, the conclusions have to be compared with empirical data and, if they are incompatible, the theories must be trashed or amended.

Many neoclassical economists have taken a principle-theory-type approach but have not been aware of some serious pitfalls. Here we again take the BS model as an example. The model takes some plausible principles which the markets are expected to follow, e.g., the no-arbitrage condition. In the meantime, however, it adopts some unrealistic assumptions, such as zero transaction cost, continuous trading, and the normal distribution of returns. More importantly, results calculated by using this model do not agree with empirical data, as illustrated in the volatility smile phenomenon described in Section 2.1.2 and the non-zero risk discussed in Section 2.3.1. Due to these flaws, it is highly doubtful whether the current application of the principle-theory-type approach in the main stream theories is valid.

The current (2008-2009) financial crisis highlights the shortcomings of standard financial theories. Buchanan (2008) pointed out that the very reason of the financial turmoil is that economists still try to understand markets by using ideas from traditional economics, especially the so-called equilibrium theory. This theory views markets as reflecting a balance of forces and changing only in response to new information, but totally neglects the internal dynamics of the markets themselves. Lohr (2008) argued that risk management models failed to keep pace with the explosive growth in complex securities. Farmer and Foley (2009) stated that the best models the policy makers have are all with fatal flaws: They assume a perfect world and by their very nature rule out crises. The policy makers are basing their decisions on common sense, but do not understand how the economy really works.