Mean-Reverting Stochastic Volatility

Jean-Pierre Fouque* George Papanicolaou† K. Ronnie Sircar‡

December 1998; revised January 1999; corrected November 1999 and January 2000.

Abstract

We present derivative pricing and estimation tools for a class of stochastic volatility models that exploit the observed "bursty" or persistent nature of stock price volatility. An empirical analysis of high-frequency S&P 500 index data confirms that volatility reverts slowly to its mean in comparison to the tick-by-tick fluctuations of the index value, but it is *fast* mean-reverting when looked at over the time scale of a derivative contract (many months). This motivates an asymptotic analysis of the partial differential equation satisfied by derivative prices, utilizing the distinction between these time scales.

The analysis yields pricing and implied volatility formulas, and the latter is used to "fit the smile" from European index option prices. The theory identifies the important group parameters that are needed for the derivative pricing and hedging problem for European-style securities, namely the average volatility and the slope and intercept of the implied volatility line, plotted as a function of the log-moneyness-to-maturity-ratio. The results considerably simplify the estimation procedure, and the data produces estimates of the three important parameters which are found to be stable within periods where the underlying volatility is close to being stationary. These segments of stationarity are identified using a wavelet-based tool.

The remaining parameters, including the growth rate of the underlying, the correlation between asset price and volatility shocks, the rate of mean-reversion of the volatility and the market price of volatility risk can be roughly estimated, but are not needed for the asymptotic pricing formulas for European derivatives. The extension to American and path-dependent contingent claims is the subject of future work.

*Department of Mathematics, North Carolina State University, Raleigh NC 27695-8205, *fouque@math.ncsu.edu*. This work was done while visiting the Department of Mathematics, Stanford University.

†Department of Mathematics, Stanford University, Stanford CA 94305, *papanico@math.stanford.edu*. Work supported by NSF grant DMS-9622854.

‡Department of Mathematics, University of Michigan, Ann Arbor MI 48109-1109, *sircar@umich.edu*. Work partially supported by University of Michigan Rackham Grant and NSF grant DMS-9803169.

Contents

1 Introduction

A derivative pricing theory is successful if the parameters that describe it remain constant when they are estimated from updated segments of historical data. Often only the simplest models have sufficient ease of tractability that the latter issue can be tested without a highly

computationally-intensive empirical study appearing years after the model is proposed. For example, the Black-Scholes theory has been of great use historically in markets and over time frames where the volatility has been close to constant.

We present here a framework for derivative pricing that is tractable enough that the stability of the parameters it needs can be investigated efficiently on large datasets that are increasingly available, and we do so with high-frequency S&P 500 index values and option prices. Such efficiency is obtained through simple asymptotic formulas that approximate the model-implied volatility surface when volatility persists, as it has been widely observed to do. Volatility clustering has not previously been used to simplify the basic pricing and estimation problems, and the methodology detailed here has many other applications to risk management and portfolio selection questions.

1.1 Background

Stochastic volatility models have become popular for derivative pricing and hedging in the last ten years as the existence of a nonflat implied volatility surface (or term-structure) has been noticed and become more pronounced, especially since the 1987 crash. This phenomenon, which is well-documented in, for example, [23, 33], stands in empirical contradiction to the consistent use of a classical Black-Scholes (constant volatility) approach to pricing options and similar securities. However, it is clearly desirable to maintain as many of the features as possible that have contributed to this model's popularity and longevity, and the natural extension pursued in the literature and in practice has been to modify the specification of volatility in the stochastic dynamics of the underlying asset price model.

Any extended model must also specify what data it is to be calibrated from. The pure Black-Scholes procedure of estimating from historical stock data *only* is not possible in an incomplete market if one takes the view (as we shall) that the market selects a unique derivative pricing measure, from a family of possible measures, which reflects its degree of "crash-o-phobia". Thus at least *some* derivative data has to be used to price other derivatives, and much recent work uses *only* derivative data to estimate all the model parameters so that the assumed relationship between the dynamics of derivative prices and the dynamics of the underlying is not exploited at all.

This is largely the case in the implied deterministic volatility (IDV) literature where volatility is modeled as a deterministic function of the asset price X_t: volatility $= \sigma(t, X_t)$. The stochastic differential equation modeling the asset price is

$$dX_t = \mu X_t dt + \sigma(t, X_t) X_t dW_t,$$

and the function $C(t, x)$ giving the no-arbitrage price of a European derivative security at time t when the asset price $X_t = x$ then satisfies the generalized Black-Scholes PDE

$$C_t + \frac{1}{2}\sigma^2(t, x)x^2 C_{xx} + r(xC_x - C) = 0,$$

with r the constant risk free interest rate and terminal condition appropriate for the contract. This has the nice feature that the market is complete which, in this context, means that the derivative's risk can (theoretically) be perfectly hedged by the underlying, and there is no volatility risk premium to be estimated.

Numerically inferred local volatility surfaces from market data by tree methods [34] or relative-entropy minimization [3] or interpolation [36] have yielded interesting qualitative properties of the (risk-neutral) probability distribution used by the market to price derivatives (such as excess skew and leptokurtosis in comparison to the lognormal distribution). In addition, these estimates are extremely useful for contemporaneous calibration of exotic securities, but this approach has not yet produced a stable surface that can be used consistently and with confidence over time. See [13] for a detailed empirical study of this issue and [25] for a mathematical explanation of why these surface-fits are outperformed by "fixed smile" (projected) implied volatilities. Possibly this shortcoming could be improved by using historical underlying data as well, though it is not clear how to implement this.

We also refer the reader to recent surveys of the stochastic volatility literature such as [16, 17, 20].

1.2 Present Approach

We concentrate on the "pure" stochastic volatility approach in which volatility σ_t is modeled as an Itô process driven by a Brownian motion that has a component independent of the Brownian motion W_t driving the asset price.

1.2.1 Stochastic Volatility World and Implied Volatility Curves

In practice, traders are given to buying and selling in units of implied volatility corresponding to option prices through the Black-Scholes formula, often known as "trading the skew". This synoptic variable has been used to express a significant discrepancy between market and Black-Scholes prices: the implied volatilities of market prices vary with strike price and time-to-maturity of the contracts. Commonly reported shapes of the curve plotted against strike price with expiration fixed, are U-shaped [33], called the smile curve and, more recently, negative or positive sloping [34], known as skew.

This particular shortcoming is remedied by stochastic volatility models first studied by Hull & White [21], Scott [35] and Wiggins [39] in 1987. The underlying asset price is modelled as a stochastic process which is now driven by a random volatility Itô process that may or may not be independent. It was shown by Renault & Touzi [32] that stochastic volatility European option prices produce the smile curve for *any* volatility process uncorrelated with the Brownian motion driving the price process, and this robustness to specific modeling of the volatility gives this extension of Black-Scholes a little more tractability than earlier ones. Of course a smile curve exhibited by options data does not necessarily imply stochastic volatility.

When there is correlation between volatility and price shocks, a similar global result is not known. However, numerical simulations in [21] with volatility a geometric Brownian motion give a negative skew for negative correlation and positive skew for positive correlation. This is confirmed by small fluctuation asymptotic results in [37] for *any* Itô volatility process, and also the results of Section 3. The explicit formulas for the implied volatility curve are different in the limit of small fluctuations and in the limit of fast mean-reversion as here. See also [25] for detailed calculations in the former regime.

1.2.2 Separation of Scales

There has been much analysis of specific Itô models in the literature by numerical and analytical methods, for example [18, 35, 38], many of which have ignored correlation effects and/or the volatility risk premium for tractability. Our goal is to identify and estimate from market data the *relevant* parameters for derivative pricing, and to test their stability over time, and thus the potential usefulness of stochastic volatility models for accurately assessing market risks and pricing exotics.

What is (to our knowledge) new here in comparison with previous empirical work on stochastic volatility models is our keeping of these two factors, use of high-frequency (intraday) data, and an asymptotic simplification of option prices predicted by the model that identifies the important *groupings of the basic parameters* that determine the observed deviation of implied volatilities from historical volatility. These turn out to be easily estimated from at-the-money market option prices.

The latter exploits the separation of time-scales introduced (in this context) in [37]. It is often observed that while volatility might fluctuate considerably over the many months comprising the lifetime of an options contract, it does not do so as rapidly as the stock price itself. That is, there are periods when the volatility is high, followed by periods when it is low. Within these periods, there might be much fluctuation of the stock price (as usual), but the volatility can be considered *relatively* constant until its next "major" fluctuation. The "minor" volatility fluctuations within these periods are relatively insignificant, especially as far as option prices, which come from an average of a functional of possible paths of the volatility, are concerned.

Many authors, for example [1], have proposed nonparametric estimation of the pricing measure for derivatives. The analysis in [37] is independent of specific modeling of the volatility process, but results in bands for option prices that describe potential volatility risk in relation to its historical autocorrelation decay structure, while obviating the need to estimate the risk premium. However, the market in at- and near-the-money European options is liquid and its historical data can be used to estimate this premium[1]. We attempt this with a parsimonious model that is complex enough to reflect an important number of observed volatility features:

1. volatility is positive;

2. volatility is mean-reverting, but persists;

3. volatility shocks are negatively correlated with asset price shocks. That is, when volatility goes up, stock prices tend to go down and *vice-versa*. This is often referred to as leverage, and it at least partially accounts for a skewed distribution for the asset price that lognormal or zero-correlation stochastic volatility models do not exhibit. The skew is documented in empirical studies of historical stock prices, for example [7], and past implied volatility data [5].

[1] This was suggested to us by Darrell Duffie.

1.3 Main Result

1. When the rate of volatility mean-reversion α, defined in (7), is large (volatility persistence), the implied volatility curve from European call options is well-approximated by a straight line in the composite variable labelled the *log-moneyness-to-maturity-ratio* (LMMR)

$$\text{LMMR} := \frac{\log\left(\frac{\text{Strike Price}}{\text{Stock Price}}\right)}{\text{Time to Maturity}}.$$

That is, if C^{call} is the stochastic volatility call option price with payoff function $h(x) = (x-K)^+$, then I defined by

$$C^{\text{call}} = C_{BS}(I),$$

where C_{BS} is the Black-Scholes formula, is given by

$$I = a\frac{\log(K/x)}{(T-t)} + b + \mathcal{O}(\alpha^{-1}). \tag{1}$$

The parameters a and b are easily estimated as the slope and intercept of the linefit.

2. The price C^h of **any** other European-style derivative with terminal payoff $h(x)$, including for example binary options and barrier options, is given by

$$C^h = C_0^h + \tilde{C}_1 + \mathcal{O}(\alpha^{-1}), \tag{2}$$

where $C_0^h(\overline{\sigma})$ is the solution to the corresponding Black-Scholes problem with constant volatility $\overline{\sigma}$, and $\tilde{C}_1(t,x)$ solves

$$\mathcal{L}_{BS}(\overline{\sigma})\tilde{C}_1 = V_3 x^3 \frac{\partial^3 C_0^h}{\partial x^3} + V_2 x^2 \frac{\partial^2 C_0^h}{\partial x^2},$$

with

$$\mathcal{L}_{BS}(\overline{\sigma}) := \frac{\partial}{\partial t} + \frac{1}{2}\overline{\sigma}^2 x^2 \frac{\partial^2}{\partial x^2} + r\left(x\frac{\partial}{\partial x} - \cdot\right), \tag{3}$$

$$V_3 := -a\overline{\sigma}^3, \tag{4}$$

$$V_2 := \overline{\sigma}\left((\overline{\sigma} - b) - a(r + \frac{3}{2}\overline{\sigma}^2)\right), \tag{5}$$

and $\overline{\sigma}$ is the long-run historical asset price volatility. The terminal condition is $\tilde{C}_1(T,x) = 0$ and any boundary conditions are zero also. The example of a knock-out barrier option is computed in [15].

The table below then distinguishes the model parameters, defined in Section 2, from the parameters that are actually needed for the theory. The latter can be written as groupings of the former by the formulas given in Section 3, but for practical purposes, there is no need to do so. We pursue this in Section 6 for empirical completeness.

Model Parameters	Parameters that are needed
Growth rate of stock μ	Mean historical volatility of stock $\overline{\sigma}$
Mean-level of log volatility m	
Rate of mean-reversion of volatility α	Slope of implied volatility linefit a
Volatility of volatility β	
Correlation between shocks ρ	Intercept of implied volatility linefit b
Volatility risk premium γ	

The three parameters on the right-side of the table are easily estimated and found to be quite stable from S&P 500 data.

Outline

Section 2 describes the basic model, its motivation and how it is used to price derivatives. The asymptotic results are given in Section 3 in which the simple implied volatility surface formula is presented. Then in Section 4, we validate use of the asymptotics using S&P 500 data to quantify volatility persistence by its (large) mean-reversion rate coefficient. The implied volatility formula is fitted to near-the-money observed smirks in Section 5 and the stability of its estimated slope and intercept over different sections of the data is demonstrated. Finally for completeness, we give ballpark estimates of the correlation and the volatility risk premium in Section 6. We conclude and outline future plans for using the separation of scales methodology in Section 7.

2 Mean-Reverting Stochastic Volatility Models

2.1 Model

We analyze models in which stock prices are conditionally lognormal, and the volatility process is a positive increasing function of a mean-reverting Ornstein-Uhlenbeck (OU) process. That is,

$$\begin{aligned} \frac{dX_t}{X_t} &= \mu dt + f(Y_t) dW_t, & (6) \\ dY_t &= \alpha(m - Y_t)dt + \beta d\hat{Z}_t, & (7) \\ \hat{Z}_t &:= \rho W_t + \sqrt{1-\rho^2} Z_t, \end{aligned}$$

where W and Z are independent Brownian motions, and ρ is the correlation between price and volatility shocks, with $|\rho| < 1$.

The solution to (7) is

$$Y_t = m + (Y_0 - m)e^{-\alpha t} + \beta \int_0^t e^{-\alpha(t-s)} d\hat{Z}_s, \tag{8}$$

and, given Y_0, Y_t is Gaussian,

$$Y_t - Y_0.e^{-\alpha t} \sim \mathcal{N}\left(m\left(1 - e^{-\alpha t}\right), \nu^2\left(1 - e^{-2\alpha t}\right)\right), \tag{9}$$

where $\nu^2 := \beta^2/(2\alpha)$. Thus Y has a unique invariant distribution, namely $\mathcal{N}(m, \nu^2)$, and is a simple building-block for a large class of stochastic volatility models described by choice of $f(\cdot)$. We call these models mean-reverting because the volatility is a monotonic function of a process Y whose drift pulls it towards the mean value m. The volatility is correspondingly pulled towards *approximately* $f(m)$. We note that another suitable building-block process is when Y_t is a mean-reverting Feller (or Cox-Ingersoll-Ross or square-root) process:

$$dY_t = \alpha(m - Y_t)dt + \beta\sqrt{Y_t}d\hat{Z}_t, \tag{10}$$

and this could be analyzed similarly. However, we believe that leaving free the choice of f affords sufficient flexibility, while our subsequent pricing formulas are structurally unchanged by different choices. The simplest example, $f(y) = e^y$ was proposed by Scott [35] and was also studied by Wiggins [39]. It is related to EGARCH models by Nelson [29]; the asymptotic analysis of Section 3 for this particular case appears in [14].

Figure 1 shows the estimate S&P 500 twenty-day transition probability density (from the high-frequency data using methods described in Section 4 and 5). It is shown in comparison to the corresponding constant volatility lognormal density. The empirical density is generated by simulation of (6)-(7) using the estimated parameter values. Clearly even the Gaussian-based volatility model fattens the tails of the lognormal distribution. The negative correlation generates the asymmetrically fatter left-tail.

2.2 Fast mean reversion

It is often noted in empirical studies of stock prices that volatility is persistent or bursty - for days at a time it is high and then, for a similar length of time, it is low. However, over the lifetime of a derivative contract (a few months), there are many such periods, and looked at on this timescale, volatility is fluctuating fast, but not as fast as the rapidly changing stock price.

In terms of our model, we say that the volatility process is fast mean-reverting relative to the yearly timescale, but slow mean-reverting by the tick-tick timescale. Since the derivative pricing and hedging problems we study are posed over the former period, we shall say that volatility exhibits fast mean-reversion without explicitly mentioning the longer timescale of reference.

The rate of mean-reversion is governed by the parameter α, in annualized units of years^{-1}. In the next section, we present empirical evidence from S&P 500 data that α is in fact large and that ν^2 is a stable $\mathcal{O}(1)$ constant, so that our large-α option pricing formulas of Section 3 can be used.

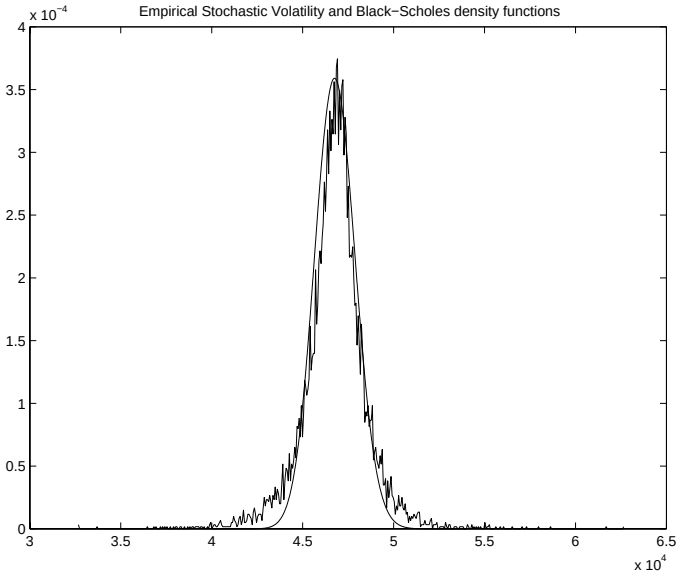

Figure 1: *Density functions for the index distribution twenty days forward. The Black-Scholes density uses the constant volatility* $\overline{\sigma}$*, and the stochastic volatility density is generated by simulation using the S&P* 500 *parameter values estimated for the first* 120 *trading days of 1994. The comparison is qualitative because of the uncertainty in these estimates, as explained in Section 6.*

As an illustration, Figure 2 shows sample stock price paths for the model (6-7) in which $\alpha = 1$ and $\alpha = 50$. Since, from (9), $\frac{1}{\alpha}\log 2$ is the time for the expected distance to the mean to halve, $\alpha = 1$ corresponds to 0.7 of a year (roughly 8 months), and $\alpha = 50$ corresponds to about half a week. Alternatively, under the invariant distribution $\mathcal{N}(m, \nu^2)$, the covariance of Y_s and Y_{s+t} is $\nu^2 e^{-\alpha t}$ and α^{-1} is the correlation time of the OU process. For $\alpha = 1$ this correlation time is a year while for $\alpha = 50$ it is about a week.

An initial visual indication that intraday S&P 500 values exhibit the kind of persistence associated with a small correlation time is shown in Figure 3, which compares the index's returns process (or normalized fluctuation sequence defined in (22)) with simulated returns processes. The data compares better with the $\alpha = 250$ simulation than the $\alpha = 1$ simulation.

2.3 Derivative Pricing

We are interested in pricing European-style derivative contracts on the underlying stock. When volatility is supposed constant, the classical Black-Scholes theory applies; when it is modeled as a stochastic process as here, the derivative price $C(t, x, y)$ is given by

$$C(t, x, y) = E^{Q(\gamma)}_{t,x,y}\{h(X_T)\}, \tag{11}$$

where $E^{Q(\gamma)}_{t,x,y}$ denotes the expectation given that $X_t = x$, $Y_t = y$, and under an Equivalent Martingale Measure (EMM) $Q(\gamma)$. The payoff function of the derivative is $h(x)$. Under such an EMM the discounted stock price is a martingale. By standard no-arbitrage pricing theory, there is more than one possible EMM because the market is incomplete (the volatility is not a traded asset); the nonuniqueness is denoted by the dependence of Q on γ, the market price

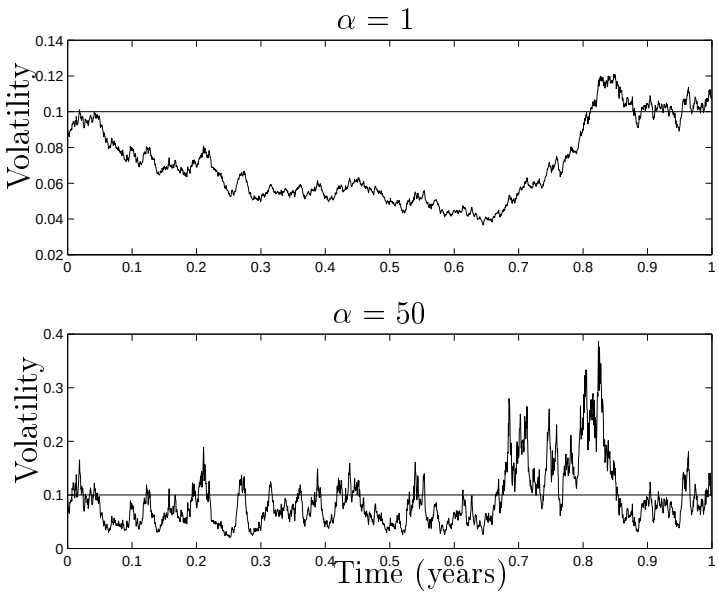

Figure 2: *Simulated paths of* $\sigma_t = f(Y_t) = e^{Y_t}$, $\{Y_t, t \geq 0\}$ *the OU process defined by (7). The top figure shows a path with* $\alpha = 1$, *and the bottom one shows a path with* $\alpha = 50$. *In both cases,* $\nu^2 = 0.25, (E\{e^{2Y_t}\})^{\frac{1}{2}} = 0.1$. *Note how volatility "clusters" in the latter case.*

of volatility risk. A detailed study of possible ways to define this concept, along with other results, is given in [25].

By Girsanov's theorem,

$$\begin{aligned} \tilde{W}_t &= W_t + \int_0^t \frac{(\mu - r)}{f(Y_s)} ds, \\ \tilde{Z}_t &= Z_t + \int_0^t \gamma_s ds, \end{aligned}$$

define independent Brownian motions $(\tilde{W}, \tilde{Z})$ under $Q(\gamma)$, assuming for instance that $(\frac{\mu - r}{f(Y_t)}, \gamma_t)$ satisfies the Novikov condition [11]. Obviously this will not be the case with $f(y) = e^y$ and Y Gaussian. Nevertheless e^y can be cutoff at 0 and the cutoff removed at the end to obtain the formula given as an example in Section 3.

The expectation in (11) is then with respect to the processes

$$\frac{dX_t}{X_t} = r dt + f(Y_t) d\tilde{W}_t, \tag{12}$$

$$dY_t = \left[\alpha(m - Y_t) - \beta\left(\rho \frac{(\mu - r)}{f(Y_t)} + \gamma_t \sqrt{1 - \rho^2}\right)\right] dt + \beta d\hat{\tilde{Z}}_t, \tag{13}$$

$$\hat{\tilde{Z}}_t := \rho \tilde{W}_t + \sqrt{1 - \rho^2} \tilde{Z}_t.$$

Further details of this derivation can be found, for example, in the review articles [16, 20]. In particular, γ_t is the risk premium factor from the *second* source of randomness Z that drives the volatility: in the perfectly correlated case $|\rho| = 1$ it does not appear, as expected. In the uncorrelated case, $\rho = 0$, γ_t is the only source of change in the drift of Y_t.

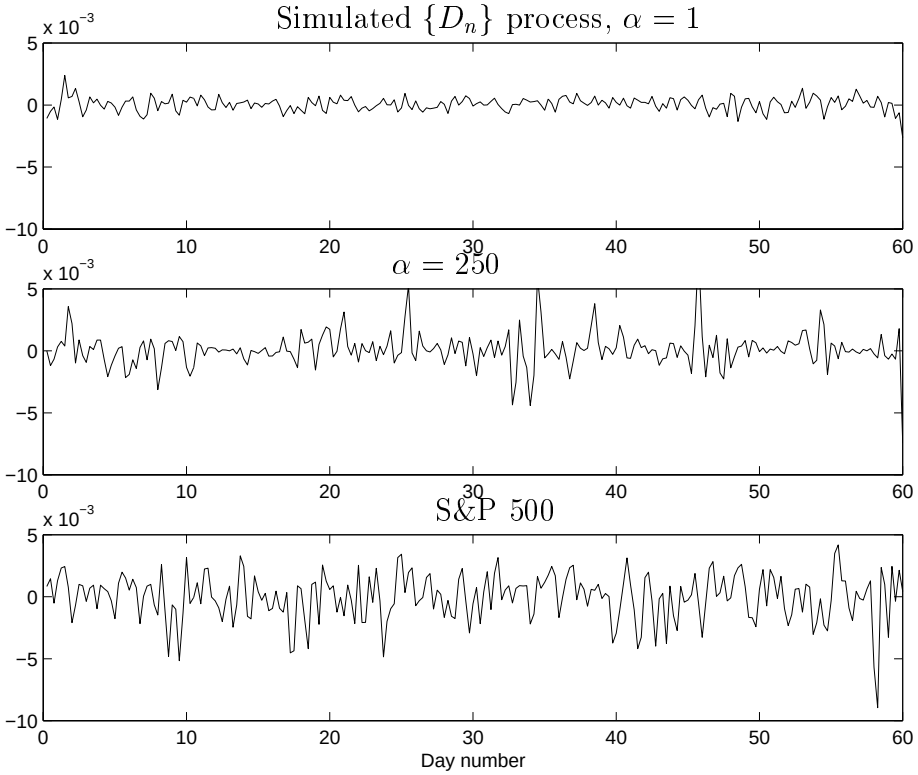

Figure 3: *The top 2 figures show simulated paths of D_n, the de-meaned returns process defined in (22) with $\alpha = 1$ and $\alpha = 250$, respectively. The other parameter values are: $\overline{\sigma} = 0.04,\ \nu = 1,\ \rho = -0.1$. The simulations are done over 60 days with 1024 points per day. This time series is then decimated to 4 points per day, leaving 240 points, which is what is shown. The 3rd figure is the S&P 500's normalized fluctuation process from the first 60 trading days of 1994 decimated to 240 points total.*

Assumption: The market price of volatility risk γ_t is constant.

As already stated, we will not need γ by itself, but rather a derived quantity containing γ that is seen in the implied volatility skew. Most studies take $\gamma = 0$ for simplicity, but we take the view that the market selects a pricing measure identified by a particular γ which will be shown to occur in a simple manner in our pricing and implied volatility formulas.

The market price of volatility risk γ may not be constant in general, just as the other parameters in the model ($\alpha,\ m,\ \beta,\ \rho,\ \mu$) might not be constant. In Section 4.2.2 we identify intervals of approximate stationarity for the historical index data wherein the model parameters can be taken as constant. The market price of volatility risk γ is not, however, determined from the historical data but from the observed option prices. We did not look for intervals of stationarity for the option prices; we simply took γ to be constant in the intervals of stationarity of the historical data. The asymptotic theory of fast mean reversion does not require constant parameter values. They can vary on the slow time scale, length $\mathcal{O}(T)$, that is, the parameters $\alpha,\ m,\ \beta,\ \rho,\ \mu$ and γ can be functions of t in (13).

In Section 3, we shall analyze the PDE corresponding to (11) in the presence of fast

mean-reversion:

$$C_t + \frac{1}{2} f(y)^2 x^2 C_{xx} + \rho\beta x f(y) C_{xy} + \frac{1}{2}\beta^2 C_{yy} + r(xC_x - C) + (\alpha(m-y) - \beta\lambda(y))\, C_y = 0, \quad (14)$$

where

$$\lambda(y) := \rho \frac{(\mu - r)}{f\,(y)} + \gamma\sqrt{1-\rho^2}. \quad (15)$$

The terminal condition is $C(T, x, y) = h(x)$.

Note In equation (13), we make the *a priori* assumption that $\gamma = \mathcal{O}(1)$: the order of the drift term in the risk-neutral volatility process is governed by α and β. From our order estimate of γ from data in Section 6, we shall *a posteriori* validate this assumption.

To summarize, the stochastic volatility model studied here is described by the five parameters $(m, \nu, \alpha, \rho, \gamma)$ which are, respectively, the mean m and the standard deviation ν of the invariant distribution of the driving OU process, the rate of mean reversion α, the skewness ρ, and the market price of volatility risk γ. The last parameter cannot be estimated from historical asset price data. As we shall see in Section 3, not all of these are needed for the pricing theory.

3 Price and Implied Volatility Formulas

Remark The results of this section do *not* assume a specific choice of $f(\cdot)$.

Now, if the rate of mean reversion α were to become larger and larger, the distinction between the time scales would disappear and the major fluctuations occur infinitely often. In this limit, volatility can be approximated by a constant as far as averages of functionals of its path are concerned (that is, weakly), and we return to the classical Black-Scholes setting. What is of interest is the next term in the asymptotic approximation of $C(t, x, y)$, valid for large α, that describes the influence of ρ, γ and the randomness ($\nu > 0$) of the volatility.

In Appendix A, we derive the following formulas for a European call option whose payoff is $h(x) = (x - K)^+$. The method of course applies to any payoff function, and there is likely to be a closed-form solution for the stochastic volatility approximations whenever there is one for the analogous classical Black-Scholes problem.

1. To lowest order, $C(t, x, y)$ is approximated by the Black-Scholes formula $C_{BS}(t, x)$ with the *OU-averaged* volatility coefficient

   $$\overline{\sigma} := \left(\langle f^2 \rangle\right)^{1/2},$$

   where $\langle \cdot \rangle$ denotes the expectation with respect to the invariant measure $\mathcal{N}(m, \nu^2)$,

   $$\langle g \rangle = \frac{1}{\sqrt{2\pi\nu^2}} \int_{-\infty}^{\infty} e^{-(y-m)^2/2\nu^2} g(y) dy.$$

   The correlation ρ and volatility risk premium γ have so far not played a role, and we only have a crude approximation around the classical theory with a suitably averaged constant volatility parameter.

2. The implied volatility surface $I(t,x;K,T)$, defined by $C(t,x,y) = C_{BS}(t,x;I)$, is correspondingly approximated at this lowest order by $\overline{\sigma}$.

3. A higher order approximation for the option price is given by

$$C(t,x,y) = C_{BS}(t,x;\overline{\sigma}) + \frac{xe^{-d_1^2/2}}{\overline{\sigma}\sqrt{2\pi}}\left(V_3\frac{d_1}{\overline{\sigma}} + (V_3 - V_2)\sqrt{T-t}\right) \tag{16}$$

where

$$\begin{aligned} V_2 &= \frac{1}{\nu\sqrt{2\alpha}}\langle[-2\rho F + \rho(\mu - r)\tilde{F} + \sqrt{1-\rho^2}\ ,\ ](f^2 - \langle f^2\rangle)\rangle, \\ V_3 &= \frac{-\rho}{\nu\sqrt{2\alpha}}\langle F(f^2 - \langle f^2\rangle)\rangle. \end{aligned}$$

d_1 is defined in (34), and $F(y)$ and $\tilde{F}(y)$ are antiderivatives of $f(y)$ and $1/f(y)$ respectively. Note that y does not explicitly appear in the first two terms of the approximation, so there is no need to estimate today's volatility.

4. The implied volatility surface is approximated (to order α^{-1}) by

$$I = a\left(\frac{\log(K/x)}{(T-t)}\right) + b + \mathcal{O}(\alpha^{-1}), \tag{17}$$

where, as in (1), we shall use the notation

$$\begin{aligned} a &= -\frac{V_3}{\bar{\sigma}^3} \\ b &= \overline{\sigma} + \frac{V_3}{\overline{\sigma}^3}\left(r + \frac{3}{2}\overline{\sigma}^2\right) - \frac{V_2}{\overline{\sigma}}. \end{aligned} \tag{18}$$

for the slope and intercept of implied volatility as a linear function of the LMMR. It remains asymptotic (and thus is a good approximation) for $\frac{T-t}{|\log(x/K)|} >> \alpha^{-1/2}$, that is, as long as the contract is not very close to expiration or very far away from the money. These extremes are not of concern here. Looked at as a surface in (K,T), the formula tells us that it is *linear* in the composite log-moneyness-to-maturity-ratio (LMMR) variable $\log(K/x)/(T-t)$, and the evolution in (t,x) is built into this variable too. This strikingly simple description is purely a feature of fast mean-reverting stochastic volatility and is independent of choice of f. A similar formula that suggested interpolation of smile/smirk curves as a functions linear in both $\log(K/x)$ and $T-t$ was derived in [37] for the case of small amplitude correlated stochastic volatility.

Note that the implied volatility curve as a function of strike price K is decreasing if $\rho < 0$ and increasing if $\rho > 0$. This ties in with numerical experiments in [21] which suggest $\text{sign}(\partial I/\partial K) = \text{sign}(\rho)$ for (in their case) lognormal volatility. The same relationship is reflected in the small fluctuation formulas for *any* correlated Itô stochastic volatility model in [37]. Zhu & Avellaneda [41] also work with a lognormal stochastic volatility and derive an explicit volatility risk premium assuming that short-term at-the-money calls are correctly

priced by Black-Scholes. Their risk-neutral volatility process has drift proportional to the correlation, and simulation reveals the same dependence of the slope of the implied volatility curve around-the-money to the correlation's sign. In addition, their large deviation asymptotics show that in the regime of large strike price (deep out-of-the-money calls), implied volatility behaves like the square root of LMMR for the model they study. That regime is not within the region of asymptoticity of Formula (17).

A typical volatility term structure predicted by this (17) (as a function of strike price and time-to-maturity at a fixed t and x) is shown in Figure 4.

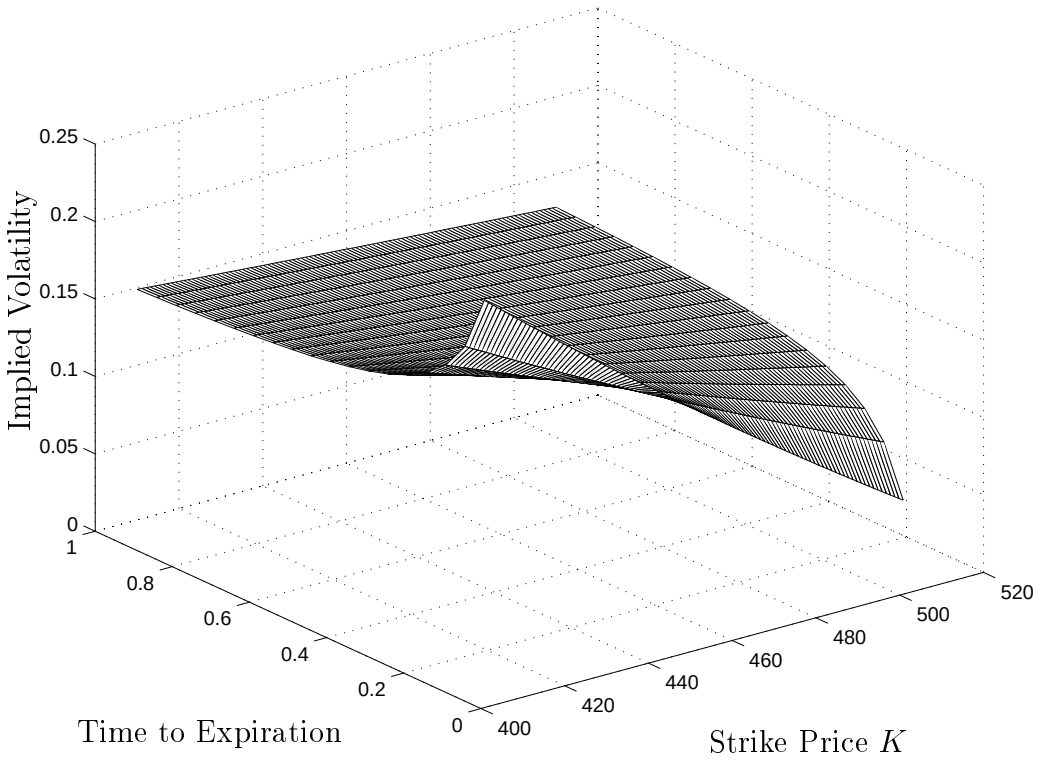

Figure 4: *Implied volatility surface from the asymptotic formula (17). The current index value is* $X_t = 460$. *The slope and intercept values of (17) are* $a = A/\sqrt{\alpha} = -0.927$ *and* $b = \overline{\sigma} + B/\sqrt{\alpha} = 0.1382$, *which are typical estimates from 1994 S&P* 500 *data found in Section 5.*

Note that m does not appear explicitly (it is contained in $\overline{\sigma}$) and that (α, ρ, γ) appear as $\rho/\sqrt{\alpha}$ and $\gamma\sqrt{1-\rho^2}/\sqrt{\alpha}$ only. In fact, the asymptotic approximation to this order for *any* European security with payoff $h(x)$ depends only on these *combined* parameters of the model for the volatility (see Appendix A for details). Thus if we obtain A and B from call option implied volatility data, the *same* values are used to price (and hedge) *all* European securities. The stability of the estimates is investigated in Section 5.

ExpOU Implied Volatility Formula

In the case $f(y) = e^y$, which we shall use in Sections 4 and 6, the averages in the constants are easily computed to give

$$I = \overline{\sigma} + (2\alpha)^{-1/2} \left(\frac{\rho}{\nu} \left(e^{3\nu^2/2} - e^{-\nu^2/2} \right) \left[\frac{\log(K/x)}{T-t} - \mu + \frac{1}{2}\overline{\sigma}^2 \right] - 2\gamma\sqrt{1-\rho^2}\nu\overline{\sigma} \right). \qquad (19)$$

For $\rho < 0$, which is the usual case, this gives a decreasing implied volatility curve when plotted against strike price K, that is, a decreasing smirk, in the exponential OU case. Note also that it is an increasing concave down function of time-to-maturity ($\sim \frac{-1}{(T-t)}$) when $K > x$, and decreasing concave up ($\sim \frac{+1}{(T-t)}$) when $K < x$.

The analysis gives rise to an explicit formula describing the geometry of the implied volatility surface across strike prices and expiration dates. In particular, the relationship to the risk premium parameter γ in (17) considerably simplifies the procedure for estimating the "crash-o-phobia" information that it contains, which otherwise would be a computationally-intensive inverse problem for the PDE (14).

Constant dividend rate It can be shown that the formulas are simply adjusted to account for a constant dividend rate $\mathcal{D}_0$ (see, for example [40, Chapter 6] for definitions). In formula (16) and in the definition of B that follows, r is replaced by $r - \mathcal{D}_0$. In the special expOU case, the $\mathcal{D}_0$ does not appear in the implied volatility formula (19).

4 Rate of Mean-Reversion of S&P 500 Volatility

Remark In the empirical work of this section, we take the model $f(y) = e^y$.

Our source of data is the Berkeley Options Database described in [22]. This gives us S&P 500 index option quoted bid-ask prices and the corresponding quoted index price, time and contract details. Since we work with quotes, the potential problem of nonsimultaneity between option and index prices does not arise. We shall present results based on European call options, taking the average of bid and ask quotes to be the current price. Strike prices are denoted by K and expiration dates by T.

Our data analysis in this work will be applied to the following dataset: the S&P 500 index and European call options on it during 1994. Looking at just the index price at different times, we have $2,340,717$ index values, starting on 3 January and ending on 30 December.

We divide the empirical work into two sections - here we use the historical index quotes only to study the speed of mean-reversion of the volatility, and in Section 5, we use option prices to obtain the slope and intercept parameters $a = A/\sqrt{\alpha}$ and $b = \overline{\sigma} + B/\sqrt{\alpha}$ of (17) which contain the market price of volatility risk, that cannot be gotten without derivatives. Only the estimate of the mean historical volatility $\overline{\sigma}$ is needed for the derivatives theory, but our study of the rate of mean-reversion will establish that it is large and that the data is within the regime of validity of the asymptotic analysis.

4.1 Review of Empirical Literature

Previous empirical work in this stochastic volatility context divides into estimation from historical stock data by moments or likelihood methods in the ARCH-related literature, fitting implied volatility alone, or "hybrid" approaches using both underlying and derivative data. An extensive review appears in [6] and we give only a brief summary.

In the first category, the EGARCH models, whose continuous-time diffusion limit is the "expOU" ($f(y) = e^y$ in (6-7)) stochastic volatility model [29], is the most relevant here because it contains skew, or nonzero correlation, whereas in ARCH/GARCH, there is none. In the original EGARCH paper [30], Nelson used maximum likelihood estimation (with a non-Gaussian random variable replacing the second Brownian increment dZ), and subsequent studies, especially in the stochastic volatility literature [8, 27, 35, 39] use the Generalized Method of Moments (GMM).

In using GMM, it is necessary to choose which moments to match and what weighting matrix to use. Indeed, there is a trade-off between a large number of moments potentially better exploiting the data, but greatly reducing the accuracy with which the weighting matrix itself can be estimated. The detailed Monte Carlo study by Andersen & Sorensen [2] of this method applied to the expOU stochastic volatility model (which they refer to as *the* stochastic volatility model) gives some guidelines in this regard, but they strongly caution against using too many moments for high-frequency data series such as ours. The empirical work mentioned so far all used daily data - for example, Scott [35] used 4 moments, while Melino & Turnbull [27] used 47. We are after specific groupings of the original parameters that the theory (Section 3) highlights as most important, so we do not undertake a global search for (α, β, m, ρ) at one attempt. In addition, we have a very large dataset whose points are non-evenly spaced - this would make the GMM procedure (especially estimation of the optimal weighting matrix) extremely complicated and computationally intensive, so we go after the parameters as we need them.

In the second category, Heynen *et al.* [19] study implied volatility as a proxy for real volatility and conclude that EGARCH models provide a better description than GARCH or CIR-based models (see equation (10)). In Merville & Piptea [28], implied volatility is found to be strongly mean-reverting, and Day & Lewis [9] find that implied volatilities contain additional information to that in historical volatilities, so that the out-of-sample predictive power of the former is greater. This motivates some authors [4, 12] to calibrate their models using derivative data only, and some "hybrid" approaches [8, 35] get most parameters from underlying data and the remaining (usually today's volatility σ_t, and/or the volatility risk premium) from least-squares fits to option prices. Here, we shall certainly use the historical index data, in the spirit of Black-Scholes, as well as near-the-money call option prices to reveal information about the market's "crash-o-phobia".

Bates [6] summarizes that most studies agree implied volatility is stationary and mean-reverting.

Note that it is difficult to directly compare numbers from previous empirical surveys such as those of Bakshi *et al.* [4], Duan [10], or Melino & Turnbull [27] because those and many others use data on the coarse daily scale for which GARCH-type models are designed. Additionally, the fitting is usually using options data only, whereas we are looking for fast mean-reversion in intraday historical data to validate the implied volatility formula (17).

Our analysis of continuous-time models assumes that time discretization is on the finest or tick-by-tick time scale and allows for the simple dependence on the group parameters. Such simplicity is not attained for coarser-grained models where estimates of all the GARCH-type parameters are needed.

4.2 Preprocessing

We use the following notation for our discrete index data. The times are $t_n, n = 0, 1, \cdots, N$, with non-uniform spacings $\Delta t_n := t_{n+1} - t_n, n = 0, 1, \cdots, N-1$. The corresponding S&P 500 index values are denoted X_n, and we consider them as realizations of the Euler discretization of (6-7):

$$\Delta X_n := X_{n+1} - X_n = X_n \left(\mu \Delta t_n + e^{Y_n} \varepsilon_n \sqrt{\Delta t_n} \right), \tag{20}$$

$$\Delta Y_n := Y_{n+1} - Y_n = \alpha(m - Y_n)\Delta t_n + \beta \hat{\eta}_n \sqrt{\Delta t_n}, \tag{21}$$

where $\hat{\eta}_n := \rho \varepsilon_n + \sqrt{(1-\rho^2)}\eta_n$, and $\{\varepsilon_n\}$ and $\{\eta_n\}$ are independent sequences of independent $\mathcal{N}(0,1)$ random variables.

We shall deal with the *normalized fluctuation sequence*:

$$D_n := \left(\frac{\Delta X_n}{X_n} - \hat{\mu} \Delta t_n \right) \frac{1}{\sqrt{\Delta t_n}}, \tag{22}$$

where

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\Delta X_n}{X_n \Delta t_n}.$$

Thus we think of D_n as a realization of $e^{Y_n} \varepsilon_n$, whose moments are easily computable[2]. When we speak of a segment of the data, the demeaning by subtraction of $\hat{\mu}$ will be done over that segment, though in practice, given our small time steps, we find that how the demeaning is done has negligible effect on the results.

4.2.1 Trading time and Subsampling

In this section, the unit of time we use is the *trading year*: this comprises the hours 9am to 3pm Central Time each trading day, with 252 trading days in the year. In other words, overnights and weekends (and holidays) are collapsed into continuous trading time. For the derivative pricing theory (and the smile-fitting of Section 5), the parameters $\overline{\sigma}, A$ and B that are needed are estimated calendar time. All we need from this section is to establish that α is large.

All our D values are computed intraday; that is, we do not compute differences that correspond to overnight differences. This gives us $2,054,462$ D_n's with between 1874 and $16,900$ per day.

In order to search for segments of stationarity and to use spectral methods for estimation, we need to subsample the data at various rates (for example to deal with vector lengths of

[2]We shall not distinguish between the random variables and their samples to keep the notation simple - which is meant will be clear in context.

powers of two that are convenient for the Fast Fourier Transform). This is done by resampling at the lower rate after filtering out high-frequencies using an eighth order Chebyshev lowpass filter. This removes the danger of aliasing that would occur from direct subsampling. We used the MATLAB process *decimate* for this purpose, and we never subsample at a rate greater than three in each use of the decimate function, and the data is subsampled from the first $2^{13}3^5 = 1,990,656$ D values, comprising just over 241 trading days. These subsampled sequences are assumed evenly spaced: this is a safe assumption because there are so many points each day. We could have linearly interpolated first, but this would not make significant difference given the high overall subsampling rate.

4.2.2 Segments of stationarity

We use the code BBLCT (Best Basis Local Cosines Transform) [26] to locate segments of the data within which the Ds can be treated as stationary. Details are given in Appendix C.

Our findings using this tool are summarized as follows:

1. In our first pass at the S&P 500 index data, we took closing prices for the five years $1993-7$ and computed 1024 daily-spaced samples of the normalized fluctuation process D. Using BBLCT, we found that one segment entirely contained the year 1994 motivating us to study the high-frequency within that year.

2. BBLCT uses a computationally expensive searching algorithm restricting its practical use to datasets of length $2^{10} = 1024$. We decimated our $1,990,656$ points down to $1,024$, subsampling successively at rates of two or three, but not more. Running BBLCT divides the 241 days into four segments of stationarity of lengths of roughly $120, 30, 30$ and 61 days. Of course the segments do not start and end exactly at day breaks, but we work on segments rounded to the nearest day.

3. In order to confirm the stability of this segmentation we go inside each of the four segments, decimate again from the full data to 1024 points per segment and run BBLCT again. We find that segments $1, 2$ and 4 are stable (that is, the routine does not re-segment it into smaller segments). Segment 3 however is quite unstable and is re-segmented into as many as ten subsegments. For now, we will take segments 1, 2 and 4 as segments of stationarity and ignore segment 3.

4.3 Estimation of $\overline{\sigma}$ and ν^2

Now that the segments of stationarity have been identified using the decimated data, we return to the high-frequency original D sequence to estimate $\overline{\sigma}$ and ν^2 within each segment.

Estimation of mean volatility

Since $D_n = e^{Y_n}\varepsilon_n$, $E\{D_n^2\} = \overline{\sigma}^2 := E\{e^{2Y_t}\}$. Thus our estimator for $\overline{\sigma}$ will be

$$\hat{\overline{\sigma}} = \sqrt{\left(\frac{1}{N}\sum_{n=0}^{N-1} D_n^2\right)}, \tag{23}$$

the square root of the sample variance of the D_n's. In the second column of Figure 5, we give the volatility estimates from the stationarity segments of the data. We will later re-estimate $\overline{\sigma}$ in calendar time for use in derivative pricing.

Segment	$\overline{\sigma}$ estimate	ν^2 estimate
1 $(1-120)$	0.0422	0.9153
2 $(121-150)$	0.0413	0.7836
4 $(181-241)$	0.0428	1.0794

Figure 5: *Estimates of* $\overline{\sigma}$ in trading time, ν^2 *and* α *from the segmented data. The length in days of each segment is given in parentheses.*

Estimating the variance of the OU process

We next estimate ν^2, the variance of the invariant measure of the OU process. Since, by direct computation in the expOU case (see Appendix B),

$$E\{D_n^4\} = 3\overline{\sigma}^4 e^{4\nu^2},$$

our estimator of ν^2 is

$$\hat{\nu^2} = \frac{1}{4}\log\left(\left(\frac{1}{N}\sum_{n=0}^{N-1} D_n^4\right)/3\hat{\overline{\sigma}}^4\right). \tag{24}$$

Estimates for each segment and subsegment are given in Figure 5. They establish that $\nu = \mathcal{O}(1)$ which provides an initial plausible basis for the stochastic volatility model we study.

Stability

To test the stability of the estimates of $\overline{\sigma}$ and ν^2 of Figure 5, we took the first segment (120 days), cut it into 49 equal pieces of $20,000$ points each and found that the fluctuation (defined by standard deviation divided by mean) of these estimators were 12.5% and 17.9% for $\hat{\overline{\sigma}}$ and $\hat{\nu^2}$ respectively.

4.4 Rate of mean reversion

Having established stable estimates of $\overline{\sigma}$ and ν^2, we now provide evidence that volatility *is* indeed fast mean-reverting, which validates applicability of the asymptotic analysis of the previous section. This is a much harder problem than obtaining the $\overline{\sigma}$ and ν^2 estimates, because it is necessary to measure correlation effects of the latent volatility process (recall that α^{-1} is like the half-life of the mean-reversion time, so we need to include lagged variables to see it). But our data is non-uniformly spaced so that equal lags in the index n do not correspond to equal lags in real time. In this section, we will subsample the data using *decimate* which removes the danger of aliasing and assume that the decimated data is equally spaced in trading time.

We present a spectral method to estimate α that indicates that α is large, or that the half-life of mean-reversion is on the order of a trading day for our data. Such a fast mean-reversion rate has previously been observed in exchange rate data [24, 27], and in equities [8].

4.4.1 Rate of mean reversion from spectra

In this section we study the rate of mean reversion α from a spectral analysis of the normalized fluctuation sequence D_n given by (22). Since for the exponential OU model that we are considering, $D_n = e^{Y_n}\varepsilon_n$, it is more convenient to do spectral analysis on $\log D_n^2 = 2Y_n + \log \varepsilon_n^2$. If our model is valid then $\log \varepsilon_n^2$ is essentially an additive white noise to the OU process whose correlation time $1/\alpha$ we want to estimate.

The spectrum of $\log D_n^2$ should be the sum of a constant background, due to the additive white noise, and a Lorenz spectrum of the form

$$\frac{\alpha\nu^2}{\alpha^2 + f^2}.$$

For large α the Lorenz spectrum will be distinguishable from the constant background for low frequencies f.

We have generated synthetic data for D_n based on the exponential OU model and carried out the spectral analysis to determine the limits of its effectiveness. We then use the spectral approach on the S&P 500 data and find that it provides striking evidence of a fast rate of mean reversion. Precise quantification of the rate is the subject of further investigation.

Synthetic data

In the simulations of the exponential OU model we use the parameters $\overline{\sigma} = 0.04$ and $\nu^2 = 1$, $\rho = -0.4$. The time length of the synthetic dataset is 120 trading days, with 32 simulated points per day. We use the Euler difference scheme (21) to obtain the sequence $\{Y_n\}$, from which we compute $D_n = e^{Y_n}\varepsilon_n$, for $n = 1, 2, \cdots, 32 \times 120 = 3840$.

We generated synthetic data with many different $\alpha's$ in the range $\alpha = 1 - 1000$. This corresponds to OU correlation times of about one year (slow mean-reversion) to an hour and a half (very fast mean-reversion). The low frequency peak seems to be present for $\alpha > 10$ and can be resolved when α is about 100 or more. We do not use data over longer periods because nonstationary effects need to be taken into consideration.

In Figure 6 we present three typical realizations of the spectrum of the logarithm of the square of a simulated normalized fluctuation process D_n, corresponding to $\alpha = 1, 10$ and 100. The horizontal frequency axis is scaled in the same units as α. In the middle and bottom graphs, we see clearly the presence of the Lorenz spectrum with an α of about 10 and 100 respectively, which can be read as the frequency where the spectrum first hits the mean-plus-5% horizontal line. A more precise criterion for identifying the α remains to be worked out. In the top graph, the Lorenz part of the spectrum in the low frequencies is not identifiable because of the lower resolution.

Similar success is seen in Figure 7 where the spectra for simulations for even larger rates of mean-reversion $\alpha = 250, 500, 1000$ are shown. Clearly the spectrum best identifies these large rates.

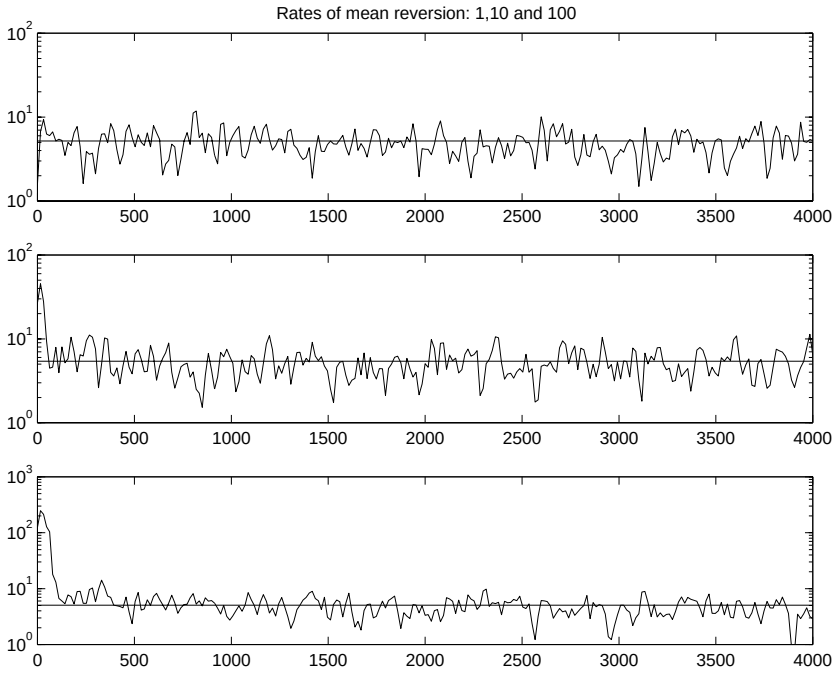

Figure 6: *Spectrum of the logarithm of the square of a* simulated *normalized fluctuation process D_n, defined in (22), with rate of mean reversion $\alpha = 1$ in the top graph, $\alpha = 10$ in the middle one and $\alpha = 100$ in the bottom. The horizontal line is the mean plus* 5%. *The data simulates* 120 *trading days with* 32 *points per day and the window size of the FFT is 512 points (*16 *days).*

4.4.2 Spectra of S&P 500 data

We went back to the original normalized fluctuation data D_n and within each of the three *stable* segments of stationarity identified by BBLCT in Figure 5, we decimated the large data set to a subsample corresponding to approximately 32 points per day. With this choice of subsampling rate, the Lorenz part of the spectrum which identifies the rate of mean reversion that much of the data exhibits (order 200) is most clearly visible.

The spectra are computed using discrete Fourier transforms on nonoverlapping windows of width *Nfft* to be chosen, and then averaged over these windows. There is a tradeoff between choosing too small a window size which results in many windows, lots of averaging and an overly-smoothed spectrum, and too big a window size which does not give enough resolution in the lower frequencies to enable detection of the Lorenz part of the spectrum.

In the long segments 1 and 4, we chose *Nfft*= 512, which then corresponds to 16 trading days on which the Lorenz part, roughly a day ($\alpha \sim 200$) can be observed. In the shorter segment 3, we take *Nfft*= 256, corresponding to 8 trading days.

In Figure 8 we show the spectra for these three segments.

Looking carefully at Figure 8, we read off the following order estimates for α in each of the stationarity segments in the manner explained above for synthetic data.

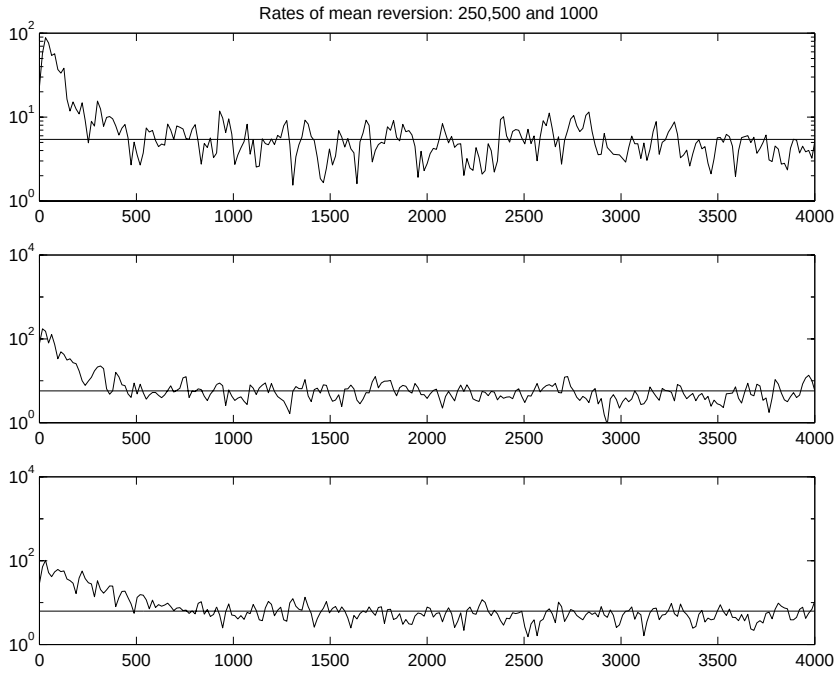

Figure 7: *Spectrum of the logarithm of the square of a* simulated *normalized fluctuation process D_n, defined in (22). From top to bottom, the rates of mean-reversion are $\alpha = 250, 500, 1000$. The data simulates* 120 *trading days with* 32 *points per day and the window size of the FFT is 512 points (*16 *days).*

Segment	α estimate (trading yrs^{-1})	Correlation time (trading days)
1 $(1-120)$	$\sim 200-250$	~ 1
2 $(121-150)$	$\sim 200-250$	~ 1
4 $(181-241)$	$\sim 200-250$	~ 1

The correlation time is 252 trading days per year/α. We note that although the estimation of α so far is rough, the rate of mean reversion is clearly large, being on the order of $200-250$.

4.4.3 Bootstrap validation of spectral method

We validate the spectral order estimate for α by simulating data using the estimated parameters $\overline{\sigma}, \nu^2$ from Figure 5 and $\alpha \sim 200$ as observed in the data in Section 4.4.2, as well as the nonuniform observation spacings $\{\Delta t_n\}$ from the real data. This is decimated in exactly the same way as we did the real data and so we can check how our subsampling method affects the computed spectrum. Then the spectra are computed using the same window and points-per-day values as for the corresponding real data, and the identical segmentation. In Figure 9, we give the results for the three segments, numbers 1, 2 and 4, and $\alpha = 200$. We chose $\rho = -0.4$ as a typical value, although we do not have estimates of this parameter yet.

These spectra compare very favorably with their corresponding real-data spectra in Figure 8. This gives us confidence that the model we are using and the subsampling and

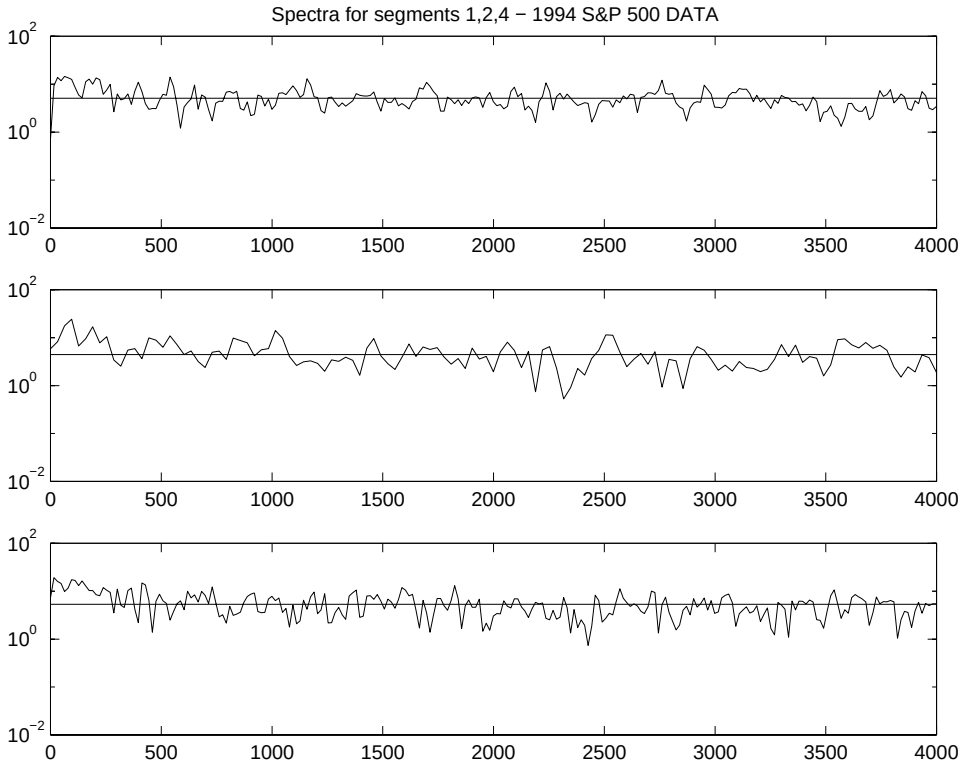

Figure 8: *Estimated spectrum of the logarithm square of the normalized fluctuation process D_n, defined in (22), for the S&P 500 in the following segments (top to bottom): segment 1 (120 trading days), segment 2 (30), segment 4 (60). The horizontal lines are the mean plus* 5%.

estimation procedure presented here capture well the mean-reverting behavior of the S&P 500 index.

4.4.4 Estimation of rate of mean reversion from time correlations

An alternative to the spectral method is to use the estimator based on the covariance of the (normalized and demeaned) squared returns:

$$E\{D_n^2 D_{n+k}^2\} \sim \overline{\sigma}^4 e^{4\nu^2 e^{-\alpha\left(t_{n+k}-t_n\right)}}, \tag{25}$$

for $\alpha\left(t_{n+k}-t_n\right) << 1$ (see Appendix B). We have also used this method on real decimated to 512 points per day, as well as simulated data.

Within each segment of stationarity, we average the empirical log-variogram given by

$$\frac{1}{4\nu^2}\log\left(\frac{\hat{E}\left\{D_n^2 D_{n+k}^2\right\}}{\hat{\overline{\sigma}}^4}\right) \tag{26}$$

in nonoverlapping subsegments. Approximating $e^{-\alpha s} \sim 1-\alpha s$ for lag time s not too large, we least-squares fit (26) to a straight line in the lag time $t_{n+k}-t_n$. The negative of the slope gives us our estimate for α. The estimates we obtain are of the same order of magnitude as the spectral estimates and have the same variability in the sense that the slope depends

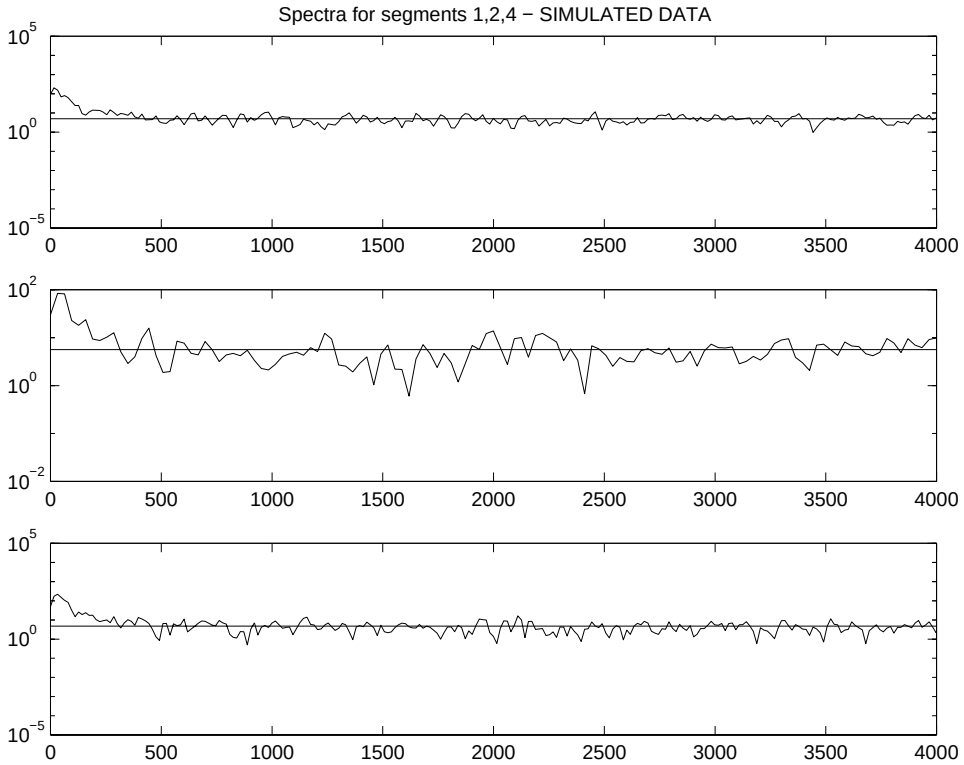

Figure 9: *Spectrum of the logarithm of the square of a* simulated *normalized fluctuation process D_n, defined in (22), with rate of mean reversion $\alpha = 200$, in the following segments (top to bottom): segment 1 (120 days), segement 2 (30), segment 4 (60). The horizontal lines are the mean plus* 5%.

sensitively on the maximum lag length chosen. This is analogous to extracting Lorenz part from the spectra. That is, the moment method confirms the presence of a *large* rate of mean reversion, but a precise estimation is not possible.

An alternative to seeking α that we also tried was to estimate the combination $\beta\rho$ using the formula (see Appendix B)

$$E\{D_n|D_{n+k}|\} = \sqrt{\frac{2}{\pi}}\beta\rho\overline{\sigma}^2\sqrt{t_{n+k}-t_n} + \mathcal{O}\left(\alpha(t_{n+k}-t_n)\right). \tag{27}$$

This moment has been used in [27, 39], for example, to obtain the correlation coefficient. However, because it appears as a higher order term for small lag times, the estimator is sensitive to the division by $\sqrt{t_{n+k}-t_n}$. Thus results can be expected to be even more highly variable, and we observed this in practice.

4.5 Remarks on estimation of the rate of mean reversion

We first summarize the essential features of the estimation procedure.

- We work in trading time.
- We identify and use segments where the data can be considered stationary.

- We extract the rate of mean-reversion from the Lorenz part of the spectrum of the logarithm of the squared normalized fluctuation process.

The positive aspects of estimation method are:

- We validate both the OU mean-reverting model and the estimation of the fast rate of mean-reversion by bootstrap.
- The method separates the intrinsic variability over segments of the model parameters from their statistical variability[3]. Note that we do not expect parameters of the volatility process to be constant across the segments of stationarity.
- Our method is well-adapted to identifying fast mean-reversion and this is what we find for the S&P 500 index.

Some negative aspects are:

- The lack of a precise quantitative estimator for the rate of mean-reversion α.
- The skew parameter ρ cannot be estimated by extension of these techniques with a comparable degree of confidence.

Neither of these two parameters is explicitly needed in practice.

5 Fitting to the S&P 500 Implied Volatility Surface

Remark Again, no specific choice of $f(\cdot)$ is assumed in this section.

We use our S&P 500 high-frequency data (described in Section 4) to fit the implied volatility surface given by (17). The unit of time is now **calendar time**, so that the time-to-expiration of a contract is measured as a fraction of the standard 365-day year.

First we take near-the-money call option prices of various strikes and maturities and compute the least-squares fit to a linear function of the LMMR variable $\log(K/x)/(T-t)$, independently of any previous estimates. This indicates initial suitability of this basis function for interpolation. We also investigate stability.

Results of smile fitting

We estimate the slope and intercept coefficients $\hat{a}$ and $\hat{b}$ from fitting Black-Scholes implied volatilities from observed option prices:

$$I^{obs}(t,x;K,T) = \hat{a}\left(\frac{\log(K/x)}{T-t}\right) + \hat{b}, \tag{28}$$

where the estimates are related to V_2, V_3 in (16) by (18).

[3] A similar analysis achieves this separation for atmospheric temperature data in [31].

We split the large first segment of stationarity (of the S&P 500 index) into four equally sized subsegments (of 30 trading days, roughly six weeks each), keep the second segment complete, and divide the fourth segment into two equally sized segments.

The results of this linear regression for each subsegment's data using only call options whose strikes are within 5% of the stock price ($|\frac{K}{x} - 1| \leq 0.05$) are given in Figure 10, and those using strikes within 3% are shown in Figure 11. We also eliminate very short-term options whose maturities are less than 3 weeks. Finally we only take options quotes from the ten-minute interval 11.10am to 11.20am each day to reduce the huge dataset which would otherwise result in overfitting of two parameters if all tick data were used. In computing the implied volatilities, we take $r = 0.048, \mathcal{D}_0 = 0.030$.

A typical distribution of strike prices (for the first half of segment 4) is shown in Figure 14, and the implied and fitted volatilities for that period in Figures 12 and 13.

Note that this is not a one-time fit of the smile or the $I(K,T)$ term structure: the evolution in (t,x) is built-in, so that we are not restricted to what option quotes are available at any given time. However, we vary t and consequently x over a small range and leave a more extensive study of the fit for future investigation.

Segment, subsegment	Total number of observations	Number of points within 5%	Slope $\hat{a}$	Intercept $\hat{b}$	St.Error ($\times 10^{-4}$)	Residual
1,1	3685	1181	-0.0578	0.1047	4.91	0.58
1,2	7272	3501	-0.0697	0.1338	1.99	0.70
1,3	10286	9041	-0.0458	0.1526	1.12	1.01
1,4	11998	9054	-0.0563	0.1403	1.71	1.55
2	3316	2251	-0.0893	0.1300	4.26	0.96
4,1	5347	2609	-0.0827	0.1430	2.55	0.67
4,2	6345	2792	-0.0829	0.1565	2.09	0.58

Figure 10: *Estimates of $\hat{a}$ and $\hat{b}$ in each segment of stationarity from call options with strikes within 5% of current index value, and more than 3 weeks to expiration. The option prices are quoted between 11.10 and 11.20 each day. St.Error denotes the standard error. The first segment is divided into 4 equal subsegments and the fourth into 2.*

We make the following observations from the results:

1. The slope coefficients $\hat{a}$ are small. This strongly supports the fast mean-reverting hypothesis and validates use of the asymptotic formula.

2. The estimates $\hat{a}$ and $\hat{b}$ within the subsegments of the stationarity segments are relatively stable (the low value of $\hat{a}$ in subsegment $1,1$ of Figure 11 seems a feature of the very low number of datapoints there). In particular, the intercept estimates are very similar in Tables 10 and 11. There is less agreement in the slopes, depending on whether 5% or 3% moneyness derivatives are used, and it remains a matter of investigation as to how close to the money data should be used before illiquidity of trading becomes a concern.

Segment, subsegment	Total number of observations	Number of points within 3%	Slope $\hat{a}$	Intercept $\hat{b}$	St.Error ($\times 10^{-4}$)	Residual
1,1	3685	453	-0.0377	0.0955	6.44	0.29
1,2	7272	1874	-0.0949	0.1333	2.87	0.54
1,3	10286	4996	-0.1191	0.1518	1.32	0.66
1,4	11998	4570	-0.0886	0.1380	2.40	1.09
2	3316	1053	-0.1270	0.1286	6.94	0.73
4,1	5347	1158	-0.0927	0.1382	3.81	0.44
4,2	6345	1186	-0.0849	0.1533	2.99	0.35

Figure 11: *Estimates of $\hat{a}$ and $\hat{b}$ in each segment of stationarity from call options with strikes within 3% of current index value, and more than 3 weeks to expiration. The option prices are quoted between 11.10 and 11.20 each day. St.Error denotes the standard error. The first segment is divided into 4 equal subsegments and the fourth into 2.*

3. The fluctuation of the slope estimates across segments is in concordance with the relative variability we found in estimates of α in Section 4 compared with the stability of the $\overline{\sigma}$ estimates: the intercept is essentially $\overline{\sigma}$ plus a correction term of order $\alpha^{-1/2}$, while the slope is an $\mathcal{O}(\alpha^{-1/2})$ term.

4. The standard error is defined as the standard deviation of the implied volatilities from the fit divided by the square root of the number of points in the fit. These errors are small indicating that $\log(K/x)/(T-t)$ is a consistent basis for interpolation. The residuals (the root of the sum of squared errors between the observations and the fit) are also very reasonable, and do not indicate that the large number of points in the sample has resulted in overfitting.

5. At the level of calibration, any European-style security can now be contemporaneously priced with $\hat{a}, \hat{b}, \overline{\sigma}$ and the asymptotic formula (2).

6 Other Parameters

Remark In this section, the empirical conclusions are for the expOU model, $f(y) = e^y$.

We now combine our fit of the implied volatility surface in Section 5 with our model-dependent estimates of $\bar{\sigma}$, ν^2 and the mean-reversion rate α from historical index data in Section 4 to obtain estimates of the skewness coefficient ρ and volatility risk premium γ for the expOU model. We shall see that while the parameters we actually need are stable, there is a large degree of uncertainty in trying to separate out the basic model parameters. The imprecision in the α and μ estimates, is inherited by estimates of ρ and γ.

Comparing (28) and (19), and knowing $\bar{\sigma}$ and ν^2, gives us the values $\rho/\sqrt{\alpha}$ from the slope $\hat{a}$ and $\gamma\sqrt{(1-\rho^2)}/\sqrt{\alpha}$ from the intercept $\hat{b}$. Using $\hat{\alpha}$ gives us the estimates

$$\hat{\rho} \;=\; \frac{\hat{a}\nu\sqrt{2\alpha}}{\left(e^{3\nu^2/2} - e^{-\nu^2/2}\right)}, \tag{29}$$

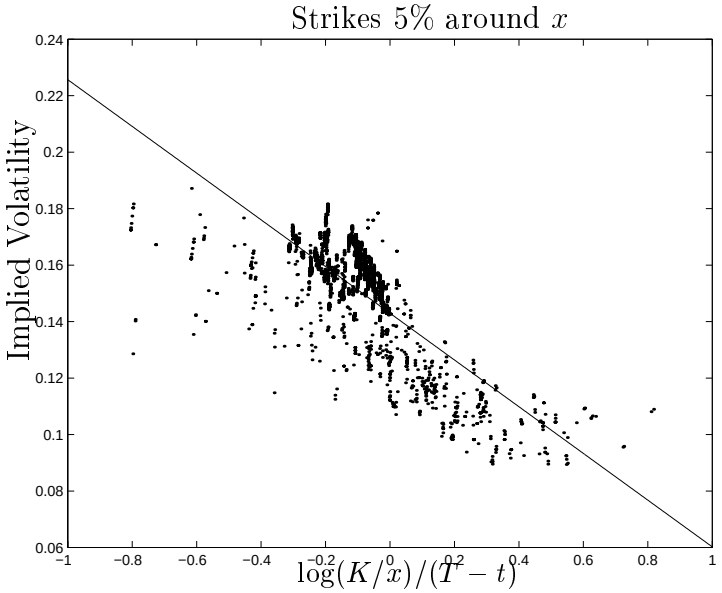

Figure 12: *Implied volatilities against log-moneyness-to-maturity-ratio (LMMR) in first half of segment* 4 *(roughly six weeks long): options with strikes within* 5% *of current index value, and more than* 3 *weeks to expiration. The straight line is the fit to formula (17).*

$$\hat{\gamma} \;=\; \frac{\hat{\overline{\sigma}} - \hat{b} - \hat{a}(\hat{\mu} - \frac{1}{2}\hat{\overline{\sigma}}^2)}{\hat{\nu}\hat{\overline{\sigma}}} \sqrt{\frac{\hat{\alpha}}{2(1-\hat{\rho}^2)}}. \tag{30}$$

Finally, we convert the estimate $\hat{\alpha}$ from Section 4.4.2 into calendar time by the factor 252/365. The conversion is of course rough because trading time is a nonlinear deformation of calendar time and not such a simple rescaling. However this inaccuracy is absorbed in the roughness of the estimate of the rate of mean-reversion. The estimate is $\overline{\sigma}$ is re-computed in calendar time using the estimator of Section 4.3 with normalized fluctuation process data $\{D_n\}$ computed in that time unit. The estimates are *not* sensitive to estimation of the mean growth rate μ.

In the table below, we compute these over the three segments of the data corresponding to those used in Table 5 and Section 4.4.2. The estimate $\hat{\mu}$ is found to be extremely sensitive to the graining of the data: high-frequency estimates are large, of the order of $1-2$, while estimates from closing prices give more familiar historical index growth rates ($|\mu| \sim \mathcal{O}(20\%)$). Nevertheless, estimates of μ are highly variable over segments and as such very unsatisfactory, but the $\hat{\mu}$ estimates do not affect the $\hat{\overline{\sigma}}$ estimates from high-frequency data because of the relative insignificance of $\mathcal{O}(\Delta t)$ terms compared to $\mathcal{O}(\sqrt{\Delta t})$ terms.

Segment (length)	$\hat{\overline{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\nu}^2$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\gamma}$
1 (6 months)	0.1015	-0.1009	0.1410	0.9153	-0.1428	∼155	∼-0.11	∼-4.97
2 ($1\frac{1}{2}$ months)	0.0994	-0.1270	0.1286	0.7835	0.4170	∼155	∼-0.20	∼2.36
4 (3 months)	0.1030	-0.0888	0.1457	1.0794	-0.0695	∼155	∼-0.065	∼-4.07

The table separates the needed parameters, whose estimates are fairly stable, from the ones presented only for completeness, whose estimates have a high degree of uncertainty.

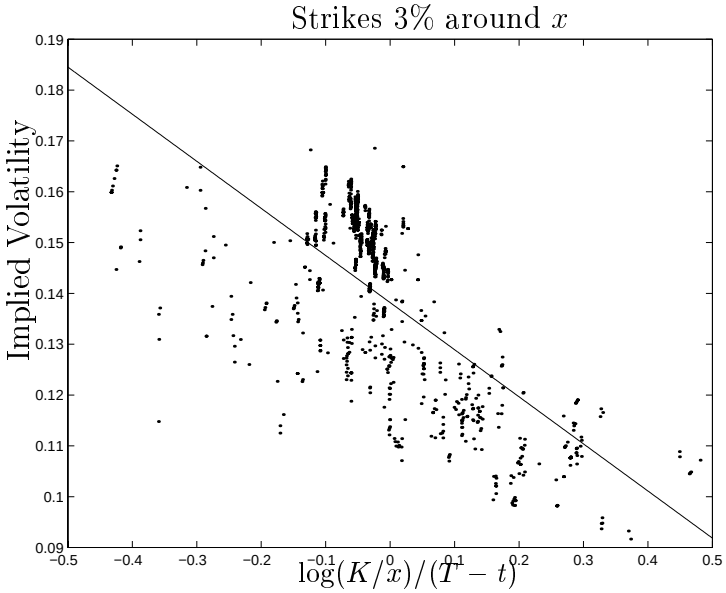

Figure 13: *Implied volatilities against log-moneyness-to-maturity-ratio (LMMR) in first half of segment* 4 *(roughly six weeks long): options with strikes within* 3% *of current index value, and more than* 3 *weeks to expiration. The straight line is the fit to formula (17).*

The data also validates *a posteriori* our assumption of γ as $\mathcal{O}(1)$ in size. That is, volatility is fast mean-reverting in the risk-neutral world as well as the real world.

7 Summary and Conclusions

1. Based on previous empirical studies and the analysis of index data presented here, volatility is well-modeled as a fast mean-reverting stochastic process. The rate of mean-reversion of the volatility is large.

2. Asymptotic analysis of the derivative pricing PDE simplifies both the forward pricing problem and the otherwise computationally-demanding inverse problem of estimating market parameters.

3. The observed S&P 500 implied volatility surface can be stably fitted to a linear function of log-moneyness-to-maturity-ratio, $\log(K/x)/(T-t)$, as suggested by the asymptotics.

4. Only the historical mean volatility and the slope and intercept of this implied volatility linefit are needed for the European pricing and hedging theory, and estimates of these from the data are stable.

5. This formula also involves in a direct way the "crash-o-phobia" information contained by the otherwise unobservable market price of volatility risk γ - it is part of the intercept of the implied volatility LMMR line.

6. Obtaining *stable* individual estimates of the risk premium itself, the stock growth rate, the correlation and the rate of mean-reversion is extremely difficult, but not necessary for the asymptotic theory.

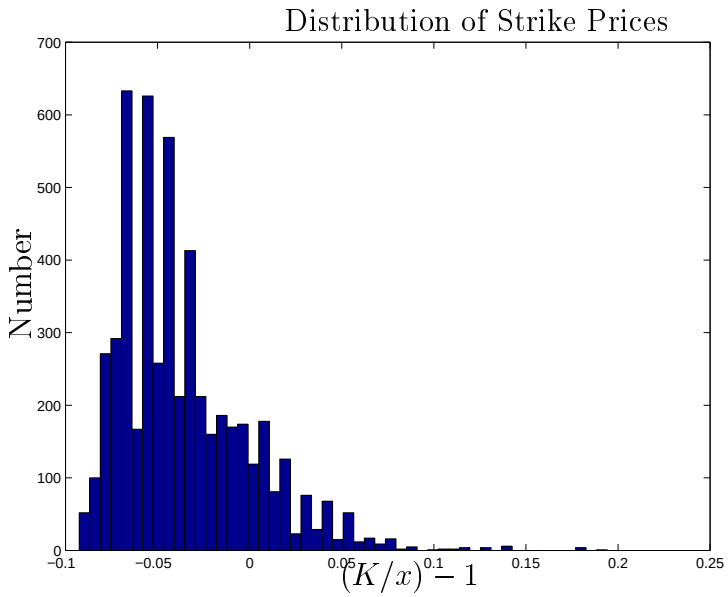

Figure 14: Distribution of call option strike prices for first half of segment 4 (roughly six weeks long). The histogram is shown in terms of moneyness $K/x - 1$.

Future directions

Empirical

The estimation tools used here can now be used to validate a fast mean-reverting model for other high-frequency datasets. We are presently preparing an empirical study of S&P 500 index data from other years, as well as foreign exchange rate data.

Fast mean-reverting methodology

1. The asymptotic expansion that gives the approximate form of the implied volatility, the smile, is general and can be used to analyze the price of other derivative instruments, for example, derivatives that depend on the price history of the underlying asset. It is necessary to go to the second term in this expansion in order to get 'smile' behavior for the implied volatility. We have constructed a more elaborate expansion in which more global information is incorporated into the first term and less in the second. But it is still necessary to have the second term for 'smile' behavior. These issues will be addressed in detail in a paper that is now in preparation.

2. We are working on an asymptotic simplification of the American option pricing problem under stochastic volatility, which currently must be solved numerically.

3. The problem of computing optimal hedging strategies under constraints when volatility is random is unsolved. For example, to optimize the probability of a successful hedge with just the underlying given an initial cash input would require solving a degenerate Hamilton-Jacobi-Bellman equation. We are looking at simplifying this problem with separation of scales asymptotics.

A Appendix: Asymptotic Analysis

We derive here the formulas presented in Section 3 and used in the empirical work of Section 5.

Let us define $\varepsilon = 1/\alpha$, so that fast mean reversion implies $0 < \varepsilon << 1$. We write (14) as $\mathcal{L}^\varepsilon C = 0$, where

$$\mathcal{L}^\varepsilon := \frac{1}{\varepsilon}\mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}}\mathcal{L}_1 + \mathcal{L}_2,$$

and

$$\begin{aligned}
\mathcal{L}_0 &:= \nu^2 \frac{\partial^2}{\partial y^2} + (m-y)\frac{\partial}{\partial y}, \\
\mathcal{L}_1 &:= \sqrt{2}\nu\rho x f(y)\frac{\partial^2}{\partial x \partial y} - \sqrt{2}\nu\lambda(y)\frac{\partial}{\partial y}, \\
\mathcal{L}_2 &:= \frac{\partial}{\partial t} + \frac{1}{2}f(y)^2 x^2 \frac{\partial^2}{\partial x^2} + r\left(x\frac{\partial}{\partial x} - \cdot\right),
\end{aligned}$$

and $\lambda(y)$ is defined in (15). Then, constructing an expansion

$$C(t,x,y) = C_0(t,x,y) + \sqrt{\varepsilon}C_1(t,x,y) + \varepsilon C_2(t,x,y) + \cdots,$$

we find, comparing powers of $\varepsilon << 1$,

$$\mathcal{L}_0 C_0 = 0$$

at the $\mathcal{O}(\varepsilon^{-1})$ level. Since $\mathcal{L}_0$ involves only y-derivatives and is the generator of the OU process Y_t, its null space is spanned by any nontrivial constant function, and it must be that C_0 does not depend on y: $C_0 = C_0(t,x)$.

At the next order, $\mathcal{O}(\varepsilon^{-1/2})$, we have

$$\mathcal{L}_1 C_0 + \mathcal{L}_0 C_1 = 0, \tag{31}$$

and since $\mathcal{L}_1$ takes y-derivatives, $\mathcal{L}_1 C_0 = 0$. By the same reasoning, (31) implies that $C_1 = C_1(t,x)$. Thus, up till $\mathcal{O}(\varepsilon)$, the option price does not depend on the current volatility.

Comparing $\mathcal{O}(1)$ terms,

$$\mathcal{L}_0 C_2 + \mathcal{L}_2 C_0 = 0.$$

Given $C_0(t,x)$, this is a Poisson equation for $C_2(t,x,y)$ and there will be no solution unless $\mathcal{L}_2 C_0$ is in the orthogonal complement of the null space of $\mathcal{L}_0^*$ (Fredholm Alternative). This is equivalent to saying that $\mathcal{L}_2 C_0$ has mean zero with respect to the invariant measure of the OU process:

$$\langle \mathcal{L}_2 C_0 \rangle = 0.$$

Since C_0 is independent of y and $\mathcal{L}_2$ only depends on y through the $f(y)$ coefficient, $\langle \mathcal{L}_2 C_0 \rangle = \langle \mathcal{L}_2 \rangle C_0$, and

$$\langle \mathcal{L}_2 \rangle = \mathcal{L}_{BS}(\overline{\sigma}) := \frac{\partial}{\partial t} + \frac{1}{2}\overline{\sigma}^2 x^2 \frac{\partial^2}{\partial x^2} + r\left(x\frac{\partial}{\partial x} - \cdot\right),$$

where $\overline{\sigma}^2 := \langle f^2 \rangle$, and $C_0(T,x) = h(x)$.

So far, we have not used the terminal condition that defines the derivative contract. To fix ideas, let us work with the European call, $h(x) = (x-K)^+$. This is used only in the computation of the terms of the asymptotic sequence, and we return to the general h problem at the end.

Thus $C_0(t,x) = C_{BS}(t,x;\overline{\sigma})$, and the first term in the expansion is the Black-Scholes (call option) pricing formula with the averaged volatility constant $\overline{\sigma}$. The γ and ρ have thus far played no role, and we proceed to find the next term in the approximation, $C_1(t,x)$.

Comparing terms of $\mathcal{O}(\sqrt{\varepsilon})$, we find

$$\mathcal{L}_0 C_3 = -\left(\mathcal{L}_1 C_2 + \mathcal{L}_2 C_1\right), \tag{32}$$

which we look at as a Poisson equation for $C_3(t,x,y)$. Just as the Fredholm solvability condition for C_2 determined the equation for C_0, the solvability for (32) will give us the equation for $C_1(t,x)$. Substituting for $C_2(t,x,y)$ with

$$C_2 = -\mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle\right) C_0,$$

this condition is

$$\left\langle \mathcal{L}_2 C_1 - \mathcal{L}_1 \mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle\right) C_0 \right\rangle = 0,$$

where

$$\langle \mathcal{L}_2 C_1 \rangle = \langle \mathcal{L}_2 \rangle C_1 = \mathcal{L}_{BS}(\overline{\sigma}) C_1$$

since C_1 does not depend on y.

Defining

$$\mathcal{A} := \left\langle \mathcal{L}_1 \mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle\right) \right\rangle,$$

the equation determining C_1 is

$$\mathcal{L}_{BS}(\overline{\sigma}) C_1 = \mathcal{A} C_0, \tag{33}$$

as C_0 does not depend on y.

Again, using that $\mathcal{L}_0$ acts only on y-dependent functions, we can write

$$\mathcal{A} = \left\langle \left(\sqrt{2}\rho\nu f(y) x \frac{\partial^2}{\partial x \partial y} - \sqrt{2}\nu\lambda(y)\frac{\partial}{\partial y}\right)\left(\frac{1}{2}\phi(y) x^2 \frac{\partial^2}{\partial x^2}\right)\right\rangle,$$

where

$$\mathcal{L}_0 \phi(y) = \nu^2 \phi''(y) + (m-y)\phi'(y) = f(y)^2 - \langle f^2 \rangle,$$

and so

$$\mathcal{A} = V x^3 \frac{\partial^3}{\partial x^3} + W x^2 \frac{\partial^2}{\partial x^2},$$

with

$$\begin{aligned} V &:= \frac{\rho\nu}{\sqrt{2}} \langle f \phi' \rangle \\ W &:= \frac{\nu}{\sqrt{2}} \langle \phi' \left(2\rho f - \lambda\right) \rangle. \end{aligned}$$

Thus we must solve

$$\begin{aligned}\mathcal{L}_{BS}(\overline{\sigma})C_1 &= Vx^3\frac{\partial^3 C_{BS}}{\partial x^3}(\overline{\sigma}) + Wx^2\frac{\partial^2 C_{BS}}{\partial x^2}(\overline{\sigma}) \\ &= \frac{xe^{-d_1^2/2}}{\overline{\sigma}\sqrt{2\pi\,(T-t)}}\left(W - V\left[1 + \frac{d_1}{\overline{\sigma}\sqrt{T-t}}\right]\right),\end{aligned}$$

where

$$d_1 = \frac{\log(x/K) + (r + \frac{1}{2}\overline{\sigma}^2)(T-t)}{\overline{\sigma}\sqrt{T-t}}, \tag{34}$$

and where we have used the explicit expression for the Black-Scholes price $C_{BS}(\overline{\sigma})$. The terminal condition is $C_1(T,x) = 0$. The explicit solution is given by the right-hand side of the PDE multiplied by $-(T-t)$. That is,

$$C_1 = \frac{xe^{-d_1^2/2}}{\overline{\sigma}\sqrt{2\pi}}\left(V\frac{d_1}{\overline{\sigma}} + (V-W)\sqrt{T-t}\right). \tag{35}$$

Finally, we compute

$$\begin{aligned}\langle f\phi'\rangle &= -\frac{1}{\nu^2}\langle F\left(f^2 - \langle f^2\rangle\right)\rangle, \\ \langle \phi'\rangle &= -\frac{1}{\nu^2}\langle y\left(f^2 - \langle f^2\rangle\right)\rangle, \\ \langle \frac{\phi'}{f}\rangle &= -\frac{1}{\nu^2}\langle \tilde{F}\left(f^2 - \langle f^2\rangle\right)\rangle,\end{aligned}$$

where $F' = f$ and $\tilde{F}' = 1/f$. Rearranging (35) gives (16), where we have labelled

$$\begin{aligned}V_2 &= W/\sqrt{\alpha} \\ V_3 &= V/\sqrt{\alpha}\end{aligned}$$

We can now calculate the implied volatility I defined by $C^\varepsilon = C_{BS}(I)$. Constructing an expansion $I = \overline{\sigma} + \sqrt{\varepsilon}I_1 + \cdots$, and comparing powers of ε, we find that

$$I_1 = C_1(t,x)\left[\frac{\partial C_{BS}}{\partial \sigma}(t,x;\overline{\sigma})\right]^{-1}, \tag{36}$$

which leads to (17), where we have expressed the expansion for I in terms of inverse of the fast mean reversion rate $\alpha = \varepsilon^{-1}$. This last computation is valid so long as $\frac{\partial C_{BS}}{\partial \sigma}$ is not very small, which only occurs as $x \to 0, \infty$ or $t \to T$.

Extension to any European derivative

Note that the procedure is identical for any European derivative with payoff $C(T,x) = h(x)$. In particular, $C_0(t,x) = C_{BS}^h(t,x;\overline{\sigma})$, where C_{BS}^h denotes the Black-Scholes price with the payoff function h, and C_1 satisfies the same PDE (33), with the same operator $\mathcal{A}$. Thus the correction term C_1 depends only on the parameters A and B (which contain the groupings $\rho/\sqrt{\alpha}$ and $\gamma\sqrt{1-\rho^2}/\sqrt{\alpha}$). It remains to be proven that this is also true for American options.

B Appendix: Estimators for the OU model parameters

We derive estimators in the continuous time situation and indicate their corresponding values in the discrete time case.

Model and parameters

We consider the case $f(y) = e^y$ and recall that our model is

$$\left\{ \begin{array}{ll} dX_t & = \mu X_t dt + e^{Y_t} X_t dW_t \\ dY_t & = \alpha(m - Y_t)dt + \beta dZ_t, \end{array} \right. \tag{37}$$

where $Z_t = \rho W_t + \sqrt{1-\rho^2}\eta_t$ for W_t and η_t two independent standard Brownian motions and $|\rho| \leq 1$. The process Y_t is the OU process under its invariant distribution $\mathcal{N}(m, \beta^2/2\alpha)$.

The process $\log X_t$ satisfies

$$d \log X_t = e^{Y_t} dW_t + (\mu - \frac{1}{2} e^{2Y_t})dt. \tag{38}$$

Our problem is to estimate the four parameters (m, α, β, ρ). Actually the useful parameters are $\bar{\sigma} = \sqrt{E\{e^{2Y_t}\}}$, $\nu^2 = \beta^2/2\alpha$, α and ρ where one can easily derive $\bar{\sigma} = e^{m+\nu^2}$. The first two parameters, determining the one-point invariant distribution, are "easy" to estimate.

$\bar{\sigma}$ estimator

For $\tau \ll 1$, at the order $\mathcal{O}(\tau)$, $E\{(\log X_{t+\tau} - \log X_t)^2\}$ depends only on the martingale term $e^{Y_t} dW_t$ in (38). Defining the martingale $M_t = \int_0^t e^{Y_s} dW_s$, $E\{(\log X_{t+\tau} - \log X_t)^2\}$ and $E\{(M_{t+\tau} - M_t)^2\}$ are equivalent at the order $\mathcal{O}(\tau)$ and

$$E\{(M_{t+\tau} - M_t)^2\} = \int_t^{t+\tau} E\{e^{2Y_s}\} ds = \tau \bar{\sigma}^2, \tag{39}$$

which gives the estimator $\hat{\bar{\sigma}}$ used in Section 4.3, (23), in the corresponding discrete situation.

ν^2 estimator

With the same notations, $E\{(\log X_{t+\tau} - \log X_t)^4\}$ and $E\{(M_{t+\tau} - M_t)^4\}$ are equivalent at the order $\mathcal{O}(\tau^2)$. This is equal to $E\{(\int_t^{t+\tau} e^{Y_s} dW_s)^4\}$ where e^{Y_s} can be replaced by e^{Y_t} at the order $\mathcal{O}(\tau^2)$. One can compute

$$E\{e^{4Y_t}(W_{t+\tau} - W_t)^4\} = E\{e^{4Y_t}\} E\{(W_{t+\tau} - W_t)^4\} = 3\tau^2 \bar{\sigma}^4 e^{4\nu^2}, \tag{40}$$

which gives the estimator $\hat{\nu^2}$ used in Section 4.3, (24), in the discrete case.

α and ρ estimators

We are now looking for autocorrelations of the squares of increments

$$E\{(\log X_{t_1+\tau} - \log X_{t_1})^2 (\log X_{t_2+\tau} - \log X_{t_2})^2\} \tag{41}$$

where we assume that

$$\tau \ll t_2 - t_1 \ll \frac{1}{\alpha}, \tag{42}$$

the last quantity, the inverse of the rate of mean reversion, being possibly small which would corresponds to our fast rate of mean reversion regime.

Again, at the order $\mathcal{O}(\tau^2)$ only the martingale term $e^{Y_t} dW_t$ in $d \log X_t$, (38), plays a role and (41) is equivalent to

$$\begin{aligned} E\left\{ (M_{t_1+\tau} - M_{t_1})^2 (M_{t_2+\tau} - M_{t_2})^2 \right\} = \\ E\left\{ \left(\int_{t_1}^{t_1+\tau} e^{Y_{s_1}} dW_{s_1} \right)^2 \left(\int_{t_2}^{t_2+\tau} e^{Y_{s_2}} dW_{s_2} \right)^2 \right\} \end{aligned} \tag{43}$$

At the order $\mathcal{O}(\tau^2)$, $e^{Y_{s_2}}$ can be replaced by $e^{Y_{t_2}}$ and $e^{Y_{s_1}}$ by $e^{Y_{t_1}}$ since $0 \leq s_2 - t_2 \leq \tau$ and $0 \leq s_1 - t_1 \leq \tau$. Using $t_1 + \tau < t_2$, one obtains that the quantity (41) is equivalent to

$$\tau E\left\{ e^{2Y_{t_2}} e^{2Y_{t_1}} (W_{t_1+\tau} - W_{t_1})^2 \right\}. \tag{44}$$

The vector $(Y_{t_1}, Y_{t_2}, W_{t_1+\tau} - W_{t_1})$ is Gaussian with mean $(m, m, 0)$ and covariance matrix

$$\begin{pmatrix} \frac{\beta^2}{2\alpha} & \frac{\beta^2}{2\alpha} e^{-\alpha(t_2-t_1)} & 0 \\ \frac{\beta^2}{2\alpha} e^{-\alpha(t_2-t_1)} & \frac{\beta^2}{2\alpha} & C \\ 0 & C & \tau \end{pmatrix}, \tag{45}$$

where the covariance C is given by

$$C = E\{(Y_{t_2} - m)(W_{t_1+\tau} - W_{t_1})\}. \tag{46}$$

Using the explicit form $Y_{t_2} - m = e^{-\alpha t_2}(Y_0 - m) + \beta e^{-\alpha t_2} \int_0^{t_2} e^{\alpha s} dZ_s$, the independence of Y_0 and $W_{t_1+\tau} - W_{t_1}$, the definition of Z_t and the independence of the Brownian motions η_t and W_t, we get that

$$\begin{aligned} C = E\left\{ \left(\rho\beta e^{-\alpha t_2} \int_0^{t_2} e^{\alpha s} dW_s \right) (W_{t_1+\tau} - W_{t_1}) \right\} = \\ E\left\{ \left(\rho\beta e^{-\alpha t_2} \int_{t_1}^{t_1+\tau} e^{\alpha s} dW_s \right) (W_{t_1+\tau} - W_{t_1}) \right\} = \\ \rho\beta e^{-\alpha t_2} \int_{t_1}^{t_1+\tau} e^{\alpha s} ds = \\ \frac{\rho\beta}{\alpha} e^{-\alpha(t_2-t_1)} (e^{\alpha\tau} - 1). \end{aligned} \tag{47}$$

Denoting by $\phi(u, v, w)$ the characteristic function of the Gaussian vector $(Y_{t_1}, Y_{t_2}, W_{t_1+\tau} - W_{t_1})$, we deduce that (44) is equal to $\tau \left(-\frac{\partial^2 \phi}{\partial w^2}(-2i, -2i, 0) \right)$ which can easily be computed to obtain that (44) is equal to

$$\tau \left(\tau + \frac{4\rho^2\beta^2}{\alpha^2} e^{-2\alpha(t_2-t_1)} (e^{\alpha\tau} - 1)^2 \right) e^{4m + \frac{2\beta^2}{\alpha}(1 + e^{-\alpha(t_2-t_1)})}. \tag{48}$$

Using the fact that $\alpha\tau \ll 1$ and keeping only the terms of order $\mathcal{O}(\tau^2)$, we get that (44), and therefore (41), is equivalent to

$$\tau^2 e^{4m+\frac{2\beta^2}{\alpha}(1+e^{-\alpha(t_2-t_1)})} = \tau^2 \bar{\sigma}^4 e^{4\nu^2 e^{-\alpha(t_2-t_1)}}, \tag{49}$$

which leads to the estimator $\hat{\alpha}$ used in Section 4.4, (26), in the discrete case. It can be seen in (48) that ρ only appears at the next order in τ. This makes it even more difficult to estimate. It is also the case for the estimator (27) and based on

$$E\{(\log X_{t_1+\tau_1} - \log X_{t_1})|\log X_{t_2+\tau_2} - \log X_{t_2}|\} \approx \sqrt{\frac{2}{\pi}}\tau_1\sqrt{\tau_2}\rho\beta\bar{\sigma}^2, \tag{50}$$

obtained similarly by using $\frac{\partial\phi}{\partial w}(-i,-i,0)$, $t_1+\tau_1 < t_2$, $\alpha\tau_1 \ll 1$ and $\alpha(t_2-t_1) \ll 1$.

C Appendix: Identification of intervals of approximate stationarity

We will describe briefly the method for identifying intervals of approximate stationarity that is presented in [26]. The software that implements this method was developed by Z. Zhang (zzhang@ms.com). It is Matlab-based and its main function is called BBLCT (best basis local cosine transform).

The spectral estimation of stationary time series is a very well developed subject. However, relatively little is known about spectral estimation when the time series is only approximately stationary. If the covariance matrix of the time series is not a Toeplitz matrix, corresponding to a stationary time series, Fourier analysis is not appropriate and the spectrum of the covariance matrix must be estimated from the data, not only the spectral coefficients. This means that the spectral analysis of a general, nonstationary, time series is quite involved and requires a very large amount of data in order to be effective. So it is natural to ask what happens when the time series is close to a stationary one so that Fourier analysis is still applicable. Another way to state this question is this: if the covariance matrix is not Toeplitz, it may still be decomposable approximately into blocks that are themselves Toeplitz, so Fourier analysis can be used in each block. This is the question that is addressed in [26].

There are a couple of issues that come up immediately: how is the segmentation into blocks to be decided and how should we measure the error in the approximation? The recently developed technology of wavelets gives us some very good tools with which to answer these questions. The main tool is the local cosine transform (LCT), which we will describe briefly in the continuous time case.

Let $a_p,\ p = 0, \pm 1, \pm 2, ...$ be a sequence of time points that go to infinity at $p \to \infty$ and negative infinity as $p \to -\infty$. The contiguous intervals $[a_p, a_{p+1}]$ form a segmentation of the time axis which we denote symbolically by γ. Given this segmentation, it is possible to construct an orthonormal basis in $L^2(R)$ which is roughly a Fourier basis in each segment. To do this we need a sequence of cutoff functions $g_p(t)$ with support that is a little bigger than the interval $[a_p, a_{p+1}]$ but smaller than the larger interval $[a_{p-1}, a_{p+2}]$, and they must

be equal to one for most of the interior of $[a_p, a_{p+1}]$. They are required to be a partition of unity $\sum_p |g_p(t)|^2 = 1$ for all t and they have to have some important even-odd properties about the interval end points $\{a_p\}$. It is the technology of wavelets that allowed the simple and explicit construction of such families of cutoff functions. Given the cutoff functions, we introduce the local Fourier transform

$$\phi_{p,k}(t) = \sqrt{\frac{2}{a_{p+1} - a_p}} g_p(t) \cos\left[\frac{(k + \frac{1}{2})\pi}{a_{p+1} - a_p}(t - a_p)\right] \tag{51}$$

where $p = 0, \pm 1, \pm 2, ...$ and $k = 0, 1, 2, ...$ The theorem of Coifman and Meyer (cited in [26] along with many other references) tells us that this collection of functions is an orthonormal basis in $L^2(R)$. To emphasize the dependence of this basis on the segmantation γ we write $\phi^{\gamma}_{p,k}(t)$.

Now given a continuous time time series X_t and a segmentation γ, we define the modulus of the Fourier coefficients of X_t relative to the basis $\phi^{\gamma}_{p,k}(t)$ by

$$d^{\gamma}_{p,k} = |\int X_t \phi^{\gamma}_{p,k}(t) dt|^2, \tag{52}$$

and the segmentation dependent functional

$$\mathcal{O}(\gamma) = \sum_{p,k} (d^{\gamma}_{p,k})^2. \tag{53}$$

This functional is just the sum of the squares of the diagonal entries of the unsmoothed, empirical covariance matrix relative to this basis. The diagonal terms $d^{\gamma}_{p,k}$ are the analog of the periodogram in the stationary case. The segmentation γ^* that maximizes this functional is such that the empirical covariance relative to the basis $\phi^{\gamma^*}_{p,k}$ is as diagonal as possible (see [26] for the detailed calculations). It is natural to think of the segmentation γ^* as the one relative to which the parts of X_t in each segment are close to stationary time series. We call the intervals corresponding to the segmentation γ^* intervals of approximate stationarity.

It is the implementation of this method that is contained in the Matlab-based software BBLCT. The details of the implementation are, naturally, complicated not only because it has to be done in discrete time and over finite segments but also because the maximization has to be done in an efficient way. Some of the implementation details are discussed in [26].

References

[1] Y. Aït-Sahalia and A. Lo. Nonparametric esimation of state-price densities implicit in financial asset prices. *J. Fin.*, 53(2), April 1998.

[2] T. Andersen and B. Sorensen. GMM estimation of a stochastic volatility model: A Monte Carlo study. *J. Bus. and Econ. Stats.*, 14(3):328–352, July 1996.

[3] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Appl. Math. Finance*, 4(1):37–64, 1997.

[4] G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *J. Fin.*, 52(5), December 1997.

[5] D. Bates. The Crash of '87: Was it expected? The evidence from options markets. *J. Fin.*, 46(3):1009–1044, July 1991.

[6] D. Bates. Testing option pricing models. In G. Maddala and C. Rao, editors, *Statistical Methods in Finance*, volume 14 of *Handbook of Statistics*, chapter 20, pages 567–611. North Holland, Amsterdam, 1996.

[7] F. Black. Studies in stock price volatility changes. In *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section*, pages 177–181. Amer. Statist. Assoc., 1976.

[8] M. Chesney and L. Scott. Pricing European Currency Options: A comparison of the modified Black-Scholes model and a random variance model. *J. Financial and Quantitative Analysis*, 24(3):267–284, September 1989.

[9] T. Day and C. Lewis. Stock market volatility and the information content of stock index options. *J. Econometrics*, 52:267–287, 1992.

[10] J-C. Duan. Cracking the Smile. *RISK*, 9(12):55–59, December 1996.

[11] D. Duffie. *Dynamic Asset Pricing Theory, 2nd ed.* Princeton University Press, 1996.

[12] D. Duffie, J. Pan, and K. Singleton. Transform analysis and option pricing for affine jump-diffusions. Technical report, Graduate School of Business, Stanford University, 1998. In preparation.

[13] B. Dumas, J. Fleming, and R. Whaley. Implied volatility functions: Empirical tests. *J. Fin.*, to appear, 1998.

[14] J.-P. Fouque, G. Papanicolaou, and K.R. Sircar. Asymptotics of a two-scale stochastic volatility model. In *Equations aux derivees partielles et applications, in honour of Jacques-Louis Lions*, pages 517–525. Gauthier-Villars, May 1998.

[15] J.-P. Fouque, G. Papanicolaou, and K.R. Sircar. Financial modeling in a fast mean-reverting stochastic volatility environment. *To appear in Columbia University Conference Proceedings*, March 1999.

[16] R. Frey. Derivative asset analysis in models with level-dependent and stochastic volatility. *CWI Quarterly*, 10(1):1–34, 1996.

[17] E. Ghysels, A. Harvey, and E. Renault. Stochastic volatility. In G. Maddala and C. Rao, editors, *Statistical Methods in Finance*, volume 14 of *Handbook of Statistics*, chapter 5, pages 119–191. North Holland, Amsterdam, 1996.

[18] 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.

[19] R. Heynen, A. Kemna, and T. Vorst. Analysis of the term strucuture of implied volatilities. *J. Fin. and Quant. Anal.*, 29(1):31–56, March 1994.

[20] D. Hobson. Stochastic volatility. Technical report, School of Mathematical Sciences, University of Bath, October 1996.

[21] J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *J. Finance*, XLII(2):281–300, June 1987.

[22] Institute of Business and Economic Research, University of California at Berkeley. *The Berkeley Options Data Base User's Guide*, June 1995. Report number 1922.

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

[24] E. Jacquier, N. Polson, and P. Rossi. Bayesian analysis of stochastic volatility models. *J. Bus. and Econ. Stats.*, 12(4):371–389, October 1994.

[25] R. Lee. Local volatilities in stochastic volatility models. *Preprint, Stanford University*, November 1998.

[26] S. Mallat, G. Papanicolaou, and Z. Zhang. Adaptive covariance estimation of locally stationary processes. *Annals of Statistics*, 26:1–47, 1998.

[27] A. Melino and S. Turnbull. Pricing foreign currency options with stochastic volatility. *J. Econometrics*, 45:239–265, 1990.

[28] L. Merville and D. Piptea. Stock-price volatility, mean-reverting diffusion, and noise. *J. Fin. Econ.*, 24:193–214, 1989.

[29] D. Nelson. Arch models as diffusion approximations. *J. Econometrics*, 45(1-2):7–38, July/Aug 1990.

[30] D. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2):347–370, March 1991.

[31] G. Papanicolaou and K. Solna. Estimation of local power law processes. *Preprint, Dept. of Mathematics, Stanford University*, August 1998.

[32] E. Renault and N. Touzi. Option Hedging and Implied Volatilities in a Stochastic Volatility Model. *Mathematical Finance*, 6(3):279–302, July 1996.

[33] M. Rubinstein. Nonparametric Tests of Alternative Option Pricing Models. *J. Finance*, XL(2):455–480, June 1985.

[34] M. Rubinstein. Implied binomial trees. *J. Finance*, LXIX(3):771–818, July 1994.

[35] L. Scott. Option Pricing when the Variance changes randomly: Theory, Estimation, and an Application. *J. Financial and Quantitative Analysis*, 22(4):419–438, December 1987.

[36] D. Shimko. Bounds of probability. *RISK*, 6(4):33–37, April 1993.

[37] K.R. Sircar and G.C. Papanicolaou. Stochastic volatility, smile and asymptotics. *Applied Mathematical Finance*, 6(1):to appear, 1999.

[38] E. Stein and J. Stein. Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, 4(4):727–752, 1991.

[39] J. Wiggins. Option Values under Stochastic Volatility. *J. Financial Economics*, 19(2):351–372, 1987.

[40] P. Wilmott, J. Dewynne, and S. Howison. *Mathematics of Financial Derivatives: A Student Introduction*. Cambridge University Press, 1996.

[41] Y. Zhu and M. Avellaneda. A risk-neutral stochastic volatility model. *Intl. J. Theor. Appl. Fin.*, 1(2):289–310, April 1998.