Maximum likelihood estimation for Heston models

Mátyás Barczy*, Mohamed Ben Alaya**,
Ahmed Kebaier**, Gyula Pap***

*University of Debrecen, **University of Paris 13, ***University of Szeged

Statistical methods for dynamical stochastic models
DYNSTOCH 2016
University Rennes 2
Rennes, 8-10 June 2016

Outline of my talk

Diffusion-type Heston model:

- definition as a special two-factor affine diffusion process
- a classification: subcritical, critical and supercritical cases (based on the asymptotic behavior of expectation vector)
- maximum likelihood estimator of the drift parameters based on continuous time observations: existence, consistency and asymptotic behavior in the above three cases.

Jump-type Heston model:

- definition of the model (only subcritical case)
- maximum likelihood estimator of the drift parameters based on continuous time observations: existence, consistency and asymptotic behaviour.

Introduction

A diffusion-type Heston process is a two-factor affine process. By diffusion-type, we mean that the sample paths are continuous almost surely.

Affine processes are common generalizations of

- continuous state and continuous time branching processes with immigration (CBI processes)

and

- Ornstein-Uhlenbeck-type (OU-type) processes.

First, we recall two-factor affine processes.

Two-factor affine processes

Definition. A time-homogeneous Markov process $(Z_t)_{t\geqslant 0}$ with state space $[0,\infty)\times\mathbb{R}$ is called a two-factor affine process if its (conditional) characteristic function takes the form

$$\mathrm{E}(\mathrm{e}^{i\langle u, Z_t\rangle} \mid Z_0 = z) = \exp\{\langle z, G(t,u)\rangle + H(t,u)\}$$

for $z\in[0,\infty)\times\mathbb{R}$, $u\in\mathbb{R}^2$ and $t\geqslant 0$, where $G(t,u)\in\mathbb{C}^2$ and $H(t,u)\in\mathbb{C}$. (Here $\langle\alpha,\beta\rangle := \alpha_1\beta_1+\alpha_2\beta_2$ for $\alpha,\beta\in\mathbb{C}^2$.)

For any $t\geqslant 0$, the (cond.) characteristic function of Z_t depends exponentially affine on the initial value z.

Duffie, Filipović and Schachermayer (2003): there exist (two-factor) affine processes for so-called admissible set of parameters.

Two-factor affine diffusion processes

Dawson and Li (2006) derived a jump-type SDE for (some) two-factor (not necessarily diffusion-type) affine processes.

We specialize this result to the diffusion case.

Theorem. (Dawson and Li (2006)) Let $a, \sigma_1, \sigma_2 > 0$, $\sigma_3 \geqslant 0$, $b, \alpha, \beta, \gamma \in \mathbb{R}$, $\varrho \in [-1, 1]$, and let us consider the SDE:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t - \gamma X_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\,\mathrm{d}(\varrho W_t + \sqrt{1-\varrho^2}B_t) + \sigma_3\,\mathrm{d}Q_t, \end{cases}$$

where $t \geqslant 0$, and $(W_t)_{t\geqslant 0}$, $(B_t)_{t\geqslant 0}$ and $(Q_t)_{t\geqslant 0}$ are independent stand. Wiener processes. Then it has a pathwise unique strong solution being a two-factor affine diffusion process.

Conversely, every two-factor affine diffusion process is a strong solution of such an SDE.

Y is a CBI process: Cox–Ingersoll–Ross (CIR) process, square root process, Feller process.

Diffusion-type Heston model (1993)

Let $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, $\varrho \in (-1, 1)$, and let us consider the SDE:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0,$$

where $(W_t)_{t\geqslant 0}$ and $(B_t)_{t\geqslant 0}$ are independent standard Wiener processes.

Google Scholar: 48 500 hits for ”Heston model”.

Aim: maximum likelihood estimators (MLEs) of the parameters a, b, α, β appearing in the drift coefficients, based on continuous time observations $(X_t)_{t\in[0,T]}$ of the process X and studying its asymptotic properties.

Overbeck (1998), Ben Alaya and Kebaier (2012, 2013):
ML estimation of (a, b) for a CIR process.

On the sample and on the parameters σ_1, σ_2 and ϱ

We observe only $(X_t)_{t\in[0,T]}$ and not $(Y_t, X_t)_{t\in[0,T]}$, where $T > 0$, since Y_t is a measurable function of $(X_s)_{s\in[0,T]}$ (i.e., a statistic) for all $t \in [0, T]$.

It is based on

$$\sum_{i=1}^{\lfloor nt \rfloor} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2 \xrightarrow{\mathrm{P}} \langle X \rangle_t = \sigma_2^2 \int_0^t Y_s \, \mathrm{d}s, \qquad \text{as } n \to \infty, \quad t \geqslant 0,$$

where $(\langle X \rangle_t)_{t\geqslant 0}$ denotes the quadratic variation process of X.

We do not estimate the parameters σ_1, σ_2 and ϱ, since these are also measurable functions (statistics) of $(X_s)_{s\in[0,T]}$.

It is based on

$$\frac{1}{\frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor} Y_{\frac{i-1}{n}}} \sum_{i=1}^{\lfloor nt \rfloor} \begin{bmatrix} Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \\ X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \end{bmatrix} \begin{bmatrix} Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \\ X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \end{bmatrix}^\top \xrightarrow{\mathrm{P}} \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

as $n \to \infty$, $t \geqslant 0$.

It turns out that for calculation of MLE of (a, b, α, β), one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

Steps of investigation

- to prove existence of a pathwise unique strong solution
- to make a classification: subcritical, critical and supercritical cases
- tools for MLE for subcritical models: ergodicity of the first coordinate process
- tools for MLE for critical models: joint Laplace transform of some functionals of Y (such as Y_t, $\int_0^t Y_s \, ds$, $\int_0^t \frac{1}{Y_s} \, ds$)
- tools for MLE for supercritical models: submartingale convergence theorems
- to derive the MLE of the drift coefficients of the model based on continuous time observations
- to study consistency and asymptotic behavior of the derived MLE as sample size tends to infinity

Existence and uniqueness of a strong solution

Proposition. Let (η_0, ζ_0) be a random vector independent of $(W_t, B_t)_{t \geqslant 0}$ satisfying $\mathrm{P}(\eta_0 \geqslant 0) = 1$. Then for $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, and $\varrho \in (-1, 1)$,

- there is a (pathwise) unique strong solution $(Y_t, X_t)_{t \geqslant 0}$ such that $\mathrm{P}((Y_0, X_0) = (\eta_0, \zeta_0)) = 1$ and $\mathrm{P}(Y_t \geqslant 0, \ \forall\, t \geqslant 0) = 1$.
- for all $s, t \geqslant 0$, $s \leqslant t$,

$$Y_t = \mathrm{e}^{-b(t-s)} \left(Y_s + a \int_s^t \mathrm{e}^{-b(s-u)} \, \mathrm{d}u + \sigma_1 \int_s^t \mathrm{e}^{-b(s-u)} \sqrt{Y_u} \, \mathrm{d}W_u \right),$$

$$X_t = X_s + \int_s^t (\alpha - \beta Y_u) \, \mathrm{d}u + \sigma_2 \int_s^t \sqrt{Y_u} \, (\varrho \, \mathrm{d}W_u + \sqrt{1 - \varrho^2} \, \mathrm{d}B_u).$$

One can check that (Y, X) is an affine process (by calculating its infinitesimal generator).

Asymptotic behavior of the expectation vector

Proposition. Let $(Y_t, X_t)_{t \geqslant 0}$ be a pathwise unique strong solution such that $\mathrm{P}(Y_0 \geqslant 0) = 1$ and $\mathrm{E}(Y_0) < \infty$, $\mathrm{E}(|X_0|) < \infty$. Then for all $t \geqslant 0$,

$$\begin{bmatrix} \mathrm{E}(Y_t) \\ \mathrm{E}(X_t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-bt} & 0 \\ -\beta \int_0^t \mathrm{e}^{-bu}\,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} \mathrm{E}(Y_0) \\ \mathrm{E}(X_0) \end{bmatrix} + \begin{bmatrix} \int_0^t \mathrm{e}^{-bu}\,\mathrm{d}u & 0 \\ -\beta \int_0^t \left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v \right) \mathrm{d}u & t \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix}.$$

Consequently, if $b > 0$, then

$$\lim_{t\to\infty} \mathrm{E}(Y_t) = \frac{a}{b}, \qquad \lim_{t\to\infty} t^{-1}\mathrm{E}(X_t) = \alpha - \frac{\beta a}{b},$$

if $b = 0$, then

$$\lim_{t\to\infty} t^{-1}\mathrm{E}(Y_t) = a, \qquad \lim_{t\to\infty} t^{-2}\mathrm{E}(X_t) = -\frac{1}{2}\beta a,$$

if $b < 0$, then

$$\lim_{t\to\infty} \mathrm{e}^{bt}\mathrm{E}(Y_t) = \mathrm{E}(Y_0) - \frac{a}{b}, \qquad \lim_{t\to\infty} \mathrm{e}^{bt}\mathrm{E}(X_t) = \frac{\beta}{b}\mathrm{E}(Y_0) - \frac{\beta a}{b^2}.$$

Classification of diffusion-type Heston processes

Based on the above asymptotic behavior of the expectation vector $[\mathrm{E}(Y_t), \mathrm{E}(X_t)]$ as $t \to \infty$, we introduce a classification of Heston processes.

Definition. Let $(Y_t, X_t)_{t \geqslant 0}$ be a unique strong solution of the given SDE satisfying $\mathrm{P}(Y_0 \geqslant 0) = 1$. We call $(Y_t, X_t)_{t \geqslant 0}$

$$\begin{cases} \text{subcritical} & \text{if } b > 0, \\ \text{critical} & \text{if } b = 0, \\ \text{supercritical} & \text{if } b < 0. \end{cases}$$

In what follows we consider Heston models with deterministic initial value $(y_0, x_0) \in (0, \infty) \times \mathbb{R}$.

Let $\mathrm{P}_{(Y,X)}$ be the probability measure on $C([0, \infty) \times \mathbb{R})$ generated by $(Y_t, X_t)_{t \geqslant 0}$, and let $\mathrm{P}_{(Y,X),T}$ be its restriction onto an appropriate σ-algebra $\mathcal{G}_T$ (past up to time T).

Existence and uniqueness of MLE, I

Lemma. Let $a \geqslant \frac{\sigma_1^2}{2}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, and $\varrho \in (-1, 1)$. Let $(Y_t, X_t)_{t \geqslant 0}$ and $(\widetilde{Y}_t, \widetilde{X}_t)_{t \geqslant 0}$ be Heston processes corresponding to the parameters $(a, b, \alpha, \beta, \sigma_1, \sigma_2, \varrho)$ and $(\sigma_1^2, 0, 0, 0, \sigma_1, \sigma_2, \varrho)$, respectively (with the same initial values). Then for all $T > 0$,

- the measures $\mathrm{P}_{(Y,X),T}$ and $\mathrm{P}_{(\widetilde{Y},\widetilde{X}),T}$ are absolutely continuous with respect to each other,
- the Radon–Nikodym derivative of $\mathrm{P}_{(Y,X),T}$ with respect to $\mathrm{P}_{(\widetilde{Y},\widetilde{X}),T}$ takes the form

$$\exp\Bigg\{ \int_0^T \frac{1}{Y_s} \begin{bmatrix} a - bY_s - \sigma_1^2 \\ \alpha - \beta Y_s \end{bmatrix}^\top \boldsymbol{S}^{-1} \begin{bmatrix} \mathrm{d}Y_s \\ \mathrm{d}X_s \end{bmatrix} - \frac{1}{2} \int_0^T \frac{1}{Y_s} \begin{bmatrix} a - bY_s - \sigma_1^2 \\ \alpha - \beta Y_s \end{bmatrix}^\top \boldsymbol{S}^{-1} \begin{bmatrix} a - bY_s + \sigma_1^2 \\ \alpha - \beta Y_s \end{bmatrix} \mathrm{d}s \Bigg\},$$

where

$$\boldsymbol{S} := \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Note that if $a \geqslant \frac{\sigma_1^2}{2}$, then $\mathrm{P}(Y_t > 0, \forall\, t \geqslant 0) = 1$.

Existence and uniqueness of MLE, II

Maximization of the loglikelihood function in $(a, b, \alpha, \beta) \in \mathbb{R}^4$ gives the MLE of (a, b, α, β) based on the sample $(X_t)_{t \in [0,T]}$:

$$\begin{bmatrix} \widehat{a}_T \\ \widehat{b}_T \\ \widehat{\alpha}_T \\ \widehat{\beta}_T \end{bmatrix} = \frac{1}{\int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{\mathrm{d}s}{Y_s} - T^2} \begin{bmatrix} \int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{\mathrm{d}Y_s}{Y_s} - T(Y_T - y_0) \\ T \int_0^T \frac{\mathrm{d}Y_s}{Y_s} - (Y_T - y_0) \int_0^T \frac{\mathrm{d}s}{Y_s} \\ \int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{\mathrm{d}X_s}{Y_s} - T(X_T - x_0) \\ T \int_0^T \frac{\mathrm{d}X_s}{Y_s} - (X_T - x_0) \int_0^T \frac{\mathrm{d}s}{Y_s} \end{bmatrix},$$

provided that $\int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{\mathrm{d}s}{Y_s} - T^2 > 0$.

We can find explicitly the global maximum point of the loglikelihood function, since, on the one hand, an explicit formula is available for the loglikelihood function, and, on the other hand, it is a quadratic expression of (a, b, α, β).

Note that for the calculation of the MLE above one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

Existence and uniqueness of MLE, III

Lemma. If $a \geqslant \frac{\sigma_1^2}{2}$, $b \in \mathbb{R}$, $\sigma_1 > 0$, and $Y_0 = y_0 > 0$, then for all $T > 0$,

$$\int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{1}{Y_s} \, \mathrm{d}s - T^2 > 0 \qquad \mathrm{P}\text{-a.s.},$$

and hence, supposing also that $\alpha, \beta \in \mathbb{R}$, $\sigma_2 > 0$, $\varrho \in (-1, 1)$, and $X_0 = x_0 \in \mathbb{R}$, there exists a unique MLE $\big(\widehat{a}_T, \widehat{b}_T, \widehat{\alpha}_T, \widehat{\beta}_T\big)$ for all $T > 0$.

Difference of the MLE and the true parameter vector

Using the given SDE, one can calculate

$$\begin{bmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \\ \widehat{\alpha}_T - \alpha \\ \widehat{\beta}_T - \beta \end{bmatrix} = \frac{1}{\int_0^T Y_s \, ds \int_0^T \frac{ds}{Y_s} - T^2} \begin{bmatrix} \sigma_1 \int_0^T Y_s \, ds \int_0^T \frac{dW_s}{\sqrt{Y_s}} - \sigma_1 T \int_0^T \sqrt{Y_s} \, dW_s \\ \sigma_1 T \int_0^T \frac{dW_s}{\sqrt{Y_s}} - \sigma_1 \int_0^T \frac{ds}{Y_s} \int_0^T \sqrt{Y_s} \, dW_s \\ \sigma_2 \int_0^T Y_s \, ds \int_0^T \frac{d\widetilde{W}_s}{\sqrt{Y_s}} - \sigma_2 T \int_0^T \sqrt{Y_s} \, d\widetilde{W}_s \\ \sigma_2 T \int_0^T \frac{d\widetilde{W}_s}{\sqrt{Y_s}} - \sigma_2 \int_0^T \frac{ds}{Y_s} \int_0^T \sqrt{Y_s} \, d\widetilde{W}_s \end{bmatrix},$$

provided that $\int_0^T Y_s \, ds \int_0^T \frac{ds}{Y_s} - T^2 > 0$, where

$$\widetilde{W}_s := \varrho W_s + \sqrt{1 - \varrho^2} B_s, \qquad s \geqslant 0,$$

is a standard Wiener process.

Consistency of MLE: subcritical case $(b > 0)$

Theorem. If $b > 0$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then

- in the case $a > \frac{\sigma_1^2}{2}$, the MLE of (a, b, α, β) is strongly consistent: $(\widehat{a}_T, \widehat{b}_T, \widehat{\alpha}_T, \widehat{\beta}_T) \xrightarrow{\text{a.s.}} (a, b, \alpha, \beta)$ as $T \to \infty$,
- in the case $a = \frac{\sigma_1^2}{2}$, the MLE of (a, b, α, β) is weakly consistent: $(\widehat{a}_T, \widehat{b}_T, \widehat{\alpha}_T, \widehat{\beta}_T) \xrightarrow{\mathrm{P}} (a, b, \alpha, \beta)$ as $T \to \infty$.

The proof is based on:

– SLLN for continuous square integrable martingales.

– If $a > \frac{\sigma_1^2}{2}$, then $\mathrm{E}(1/Y_\infty) < \infty$, and hence

$$\frac{1}{T} \int_0^T \frac{1}{Y_s} \, \mathrm{d}s \xrightarrow{\text{a.s.}} \mathrm{E}\left(\frac{1}{Y_\infty}\right), \qquad \text{as } T \to \infty,$$

where Y_∞ denotes a random variable having the unique stationary distribution of the subcritical model.

– If $a = \frac{\sigma_1^2}{2}$, then $\frac{1}{T^2} \int_0^T \frac{1}{Y_s} \, \mathrm{d}s$ converges in distribution as $T \to \infty$.

Consistency of MLE: critical and supercritical cases ($b = 0$, resp. $b < 0$)

Critical Heston models with $a > \frac{\sigma_1^2}{2}$:
the MLE of (a, b, α, β) is weakly consistent as a consequence of our forthcoming results for asymptotic behavior.

We have not investigated the case of $a = \frac{\sigma_1^2}{2}$.

Supercritical Heston models with $a \geqslant \frac{\sigma_1^2}{2}$:

- the MLE of b is strongly consistent,
- the MLE of β is weakly consistent,
- the MLEs of a and α *are not weakly consistent.*

The last three statements are consequences of our forthcoming results for asymptotic behavior.

Asymptotics of MLE: subcritical case $(b > 0)$

Theorem. If $a > \frac{\sigma_1^2}{2}$, $b > 0$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then the MLE of (a, b, α, β) is asymptotically normal, i.e.,

$$\sqrt{T} \begin{bmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \\ \widehat{\alpha}_T - \alpha \\ \widehat{\beta}_T - \beta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4 \left(\mathbf{0}, \boldsymbol{S} \otimes \begin{bmatrix} \frac{2b}{2a - \sigma_1^2} & -1 \\ -1 & \frac{a}{b} \end{bmatrix}^{-1} \right) \qquad \text{as } T \to \infty,$$

where $\otimes$ denotes tensor product of matrices and

$$\boldsymbol{S} = \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Asymptotics of MLE: critical case $(b = 0)$

Theorem. If $a > \frac{\sigma_1^2}{2}$, $b = 0$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$ and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then

$$\begin{bmatrix} \sqrt{\log T}(\widehat{a}_T - a) \\ \sqrt{\log T}(\widehat{\alpha}_T - \alpha) \\ T\widehat{b}_T \\ T(\widehat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \left(a - \frac{\sigma_1^2}{2}\right)^{1/2} \boldsymbol{S}^{1/2} \boldsymbol{Z}_2 \\ \frac{a - \mathcal{Y}_1}{\int_0^1 \mathcal{Y}_s \, \mathrm{d}s} \\ \frac{\alpha - \mathcal{X}_1}{\int_0^1 \mathcal{Y}_s \, \mathrm{d}s} \end{bmatrix} \qquad \text{as } T \to \infty,$$

where $(\mathcal{Y}_t, \mathcal{X}_t)_{t \geqslant 0}$ is a critical Heston process given by

$$\begin{cases} \mathrm{d}\mathcal{Y}_t = a \, \mathrm{d}t + \sigma_1 \sqrt{\mathcal{Y}_t} \, \mathrm{d}\mathcal{W}_t, \\ \mathrm{d}\mathcal{X}_t = \alpha \, \mathrm{d}t + \sigma_2 \sqrt{\mathcal{Y}_t} \left(\varrho \, \mathrm{d}\mathcal{W}_t + \sqrt{1 - \varrho^2} \, \mathrm{d}\mathcal{B}_t\right), \end{cases} \qquad t \geqslant 0,$$

with initial value $(\mathcal{Y}_0, \mathcal{X}_0) = (0, 0)$, where $\mathcal{W}$ and $\mathcal{B}$ are independent standard Wiener processes, $\boldsymbol{Z}_2$ is a 2-dimensional standard normally distributed random vector independent of $\left(\mathcal{Y}_1, \int_0^1 \mathcal{Y}_s \, \mathrm{d}s, \mathcal{X}_1\right)$, and $\boldsymbol{S}^{1/2}$ denotes the uniquely determined symmetric, positive definite square root of $\boldsymbol{S}$.

Asymptotics of MLE: supercritical case $(b < 0)$

Theorem. If $a \geqslant \frac{\sigma_1^2}{2}$, $b < 0$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then

$$\begin{bmatrix} \widehat{a}_T - a \\ \widehat{\alpha}_T - \alpha \\ e^{-bT/2}(\widehat{b}_T - b) \\ e^{-bT/2}(\widehat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \widetilde{\mathcal{V}} \\ \varrho \frac{\sigma_2}{\sigma_1} \widetilde{\mathcal{V}} + \sigma_2 \sqrt{1 - \varrho^2} \left(\int_0^{-1/b} \widetilde{\mathcal{Y}}_s \, ds \right)^{-1/2} Z_1 \\ \left(-\frac{\widetilde{\mathcal{Y}}_{-1/b}}{b} \right)^{-1/2} \boldsymbol{S}^{1/2} \boldsymbol{Z}_2 \end{bmatrix}$$

as $T \to \infty$, where $\widetilde{\mathcal{Y}}$ is a critical CIR process given by

$$d\widetilde{\mathcal{Y}}_t = a \, dt + \sigma_1 \sqrt{\widetilde{\mathcal{Y}}_t} \, d\mathcal{W}_t, \qquad t \geqslant 0,$$

with initial value $\widetilde{\mathcal{Y}}_0 = y_0$, where $(\mathcal{W}_t)_{t \geqslant 0}$ is a standard Wiener process,

$$\widetilde{\mathcal{V}} := \frac{\log \widetilde{\mathcal{Y}}_{-1/b} - \log y_0}{\int_0^{-1/b} \widetilde{\mathcal{Y}}_s \, ds} + \frac{\sigma_1^2}{2} - a,$$

Z_1 is a 1-dimensional stand. normally distributed random variable, $\boldsymbol{Z}_2$ is a 2-dimensional stand. normally distributed random vector such that $(\widetilde{\mathcal{Y}}_{-1/b}, \int_0^{-1/b} \widetilde{\mathcal{Y}}_s \, ds)$, Z_1 and $\boldsymbol{Z}_2$ are independent.

Diffusion-type Heston models

We considered a diffusion-type Heston model:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0,$$

where $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $\varrho \in (-1, 1)$.

In fact, the original diffusion-type Heston model takes the form

$$\begin{cases} \mathrm{d}Y_t = \kappa(\theta - Y_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}S_t = \mu S_t\,\mathrm{d}t + S_t\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0,$$

where $\kappa, \theta, \sigma > 0$, $\mu \in \mathbb{R}$, and $\varrho \in (-1, 1)$.

Differences between the two models:

- $(X_t)_{t\geqslant 0}$ is the log-price process of an asset, while $(S_t)_{t\geqslant 0}$ is the price process itself.
- the first model can be subcritical, critical or supercritical (according to $b > 0$, $b = 0$, and $b < 0$), but the second model can only be subcritical (since $\kappa > 0$).

Jump-type Heston models (SVJ models)

Let us consider the jump-type SDE

$$\begin{cases} \mathrm{d}Y_t = \kappa(\theta - Y_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}S_t = \mu S_t\,\mathrm{d}t + S_t\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big) + S_{t-}\,\mathrm{d}L_t, \end{cases} \qquad t \geqslant 0,$$

where $(L_t)_{t\geqslant 0}$ is a purely non-Gaussian Lévy process independent of $(W_t, B_t)_{t\geqslant 0}$ with Lévy–Khintchine representation

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}uL_1}) = \exp\left\{\mathrm{i}\gamma u + \int_{-1}^{\infty}(\mathrm{e}^{\mathrm{i}uz} - 1 - \mathrm{i}uz\mathbb{1}_{[-1,1]}(z))\,m(\mathrm{d}z)\right\}, \quad u \in \mathbb{R},$$

where $\gamma \in \mathbb{R}$ and m is a Lévy measure concentrating on $(-1, \infty)$ with $m(\{0\}) = 0$. (Integration with respect to L is meant via its Lévy-Itô decomposition.)

Aim: derive and study the MLE of the parameter $\psi := (\theta, \kappa, \mu)$ based on cont. time observations $(S_t, L_t)_{t\in[0,T]}$ with $T > 0$, starting the process (Y, S) from some deterministic initial value $(y_0, s_0) \in (0, \infty)^2$ supposing that σ, ϱ, γ and the Lévy measure m are known.

Jump-type Heston models: a special case

This model is quite popular in finance with the special choice of the Lévy process L as a compound Poisson process:

$$L_t := \sum_{i=1}^{\pi_t} (\mathrm{e}^{J_i} - 1), \qquad t \geqslant 0,$$

where

- $(\pi_t)_{t \geqslant 0}$ is a Poisson process with intensity 1,
- $(J_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. random variables having no atom at zero (i.e., $\mathrm{P}(J_1 = 0) = 0$),
- π, $(J_i)_{i \in \mathbb{N}}$, W and B are independent.

In this case S takes the form

$$\frac{S_t}{S_0} = \exp\left\{ \int_0^t \left(\mu - \frac{1}{2} Y_u \right) \mathrm{d}u + \int_0^t \sqrt{Y_u} \, (\varrho \, \mathrm{d}W_u + \sqrt{1 - \varrho^2} \, \mathrm{d}B_u) + \sum_{i=1}^{\pi_t} J_i \right\},$$

and hence J can be interpreted as the jump size of the logprice.

This special model has been studied, e.g., by Bakshi, Cao and Chen (1997), Broadie and Kaya (2006), Runggaldier (2003) and Sun and Guo (2015).

Existence and uniqueness of a strong solution

Proposition. Let (η_0, ζ_0) be a random vector such that η_0 is independent of $(W_t)_{t\geqslant 0}$ and $\mathrm{P}(\eta_0 \geqslant 0,\, \zeta_0 > 0) = 1$. Then for all $\theta, \kappa, \sigma \in (0, \infty)$, $\mu \in \mathbb{R}$, and $\varrho \in (-1, 1)$,

- there is a (pathwise) unique strong solution $(Y_t, S_t)_{t\geqslant 0}$ such that $\mathrm{P}((Y_0, S_0) = (\eta_0, \zeta_0)) = 1$ and $\mathrm{P}(Y_t \geqslant 0 \text{ and } S_t > 0 \text{ for all } t \geqslant 0) = 1$.
- with the notation $\Delta L_u := L_u - L_{u-}$, $u > 0$, $\Delta L_0 := 0$,

$$\frac{S_t}{S_0} = \exp\left\{ \int_0^t \left(\mu - \frac{1}{2} Y_u\right) \mathrm{d}u + \int_0^t \sqrt{Y_u}\,(\varrho\,\mathrm{d}W_u + \sqrt{1-\varrho^2}\,\mathrm{d}B_u) + L_t \right\} \times \prod_{u\in[0,t]} (1 + \Delta L_u)\mathrm{e}^{-\Delta L_u}, \qquad t \geqslant 0.$$

- if, in addition, $\theta\kappa \geqslant \frac{\sigma^2}{2}$ and $\mathrm{P}(\eta_0 > 0) = 1$, then $\mathrm{P}(Y_t > 0 \text{ for all } t \geqslant 0) = 1$.

Note that, since $\mathrm{P}(S_t > 0 \text{ for all } t \geqslant 0) = 1$, the process S can be used for modeling prices in a financial market.

Existence and uniqueness of MLE, I

From now on, we consider the jump-type Heston model with known $\sigma > 0$, $\varrho \in (-1, 1)$, $\gamma \in \mathbb{R}$, Lévy measure m, and deterministic initial value $(Y_0, S_0) = (y_0, s_0) \in (0, \infty)^2$.

We consider $\psi = (\theta, \kappa, \mu) \in (0, \infty)^2 \times \mathbb{R} =: \Psi$ as a parameter.

Let P_ψ denote the probability measure induced by $(Y_t, S_t)_{t \geqslant 0}$ on the measurable space $(D([0, \infty), \mathbb{R}^2), \mathcal{D}([0, \infty), \mathbb{R}^2))$ endowed with the natural filtration $(\mathcal{D}_t([0, \infty), \mathbb{R}^2))_{t \geqslant 0}$.

For $T > 0$, let $\mathsf{P}_{\psi,T} := \mathsf{P}_\psi|_{\mathcal{D}_T([0,\infty),\mathbb{R}^2)}$ be the restriction of P_ψ to $\mathcal{D}_T([0, \infty), \mathbb{R}^2)$.

The next lemma is about the existence and the form of the Radon–Nikodym derivative $\frac{\mathrm{d}\mathsf{P}_{\psi,T}}{\mathrm{d}\mathsf{P}_{\widetilde{\psi},T}}$ for certain $\psi, \widetilde{\psi} \in \Psi$.

Let $(\widetilde{Y}, \widetilde{S})$ denote a (reference) jump-type Heston process corresponding to $(\widetilde{\theta}, \widetilde{\kappa}, \widetilde{\mu})$ (the other parameters and the initial value are unchanged).

Existence and uniqueness of MLE, II

Lemma. Let $\psi = (\theta, \kappa, \mu) \in \Psi$ and $\widetilde{\psi} := (\widetilde{\theta}, \widetilde{\kappa}, \widetilde{\mu}) \in \Psi$ with $\theta\kappa, \widetilde{\theta}\widetilde{\kappa} \geqslant \frac{\sigma^2}{2}$. Then, for $T > 0$, the probability measures $\mathsf{P}_{\psi,T}$ and $\mathsf{P}_{\widetilde{\psi},T}$ are absolutely continuous with resp. to each other, and the log-likelihood function $\log \frac{\mathrm{d}\mathsf{P}_{\psi,T}}{\mathrm{d}\mathsf{P}_{\widetilde{\psi},T}}(\widetilde{Y}, \widetilde{S})$ takes the form

$$-\frac{1}{2}\big(H(\psi) - H(\widetilde{\psi})\big)^\top \widetilde{\boldsymbol{G}}_T \big(H(\psi) + H(\widetilde{\psi})\big) + \big(H(\psi) - H(\widetilde{\psi})\big)^\top \widetilde{\boldsymbol{f}}_T,$$

where

- $H : \mathbb{R}^3 \to \mathbb{R}^3$, $H(x_1, x_2, x_3) := (x_1 x_2, x_2, x_3)^\top$, $\boldsymbol{x} \in \mathbb{R}^3$,
-

$$\widetilde{\boldsymbol{G}}_T := \frac{1}{(1-\varrho^2)\sigma^2} \int_0^T \frac{1}{\widetilde{Y}_u} \begin{bmatrix} 1 & -\widetilde{Y}_u & -\varrho\sigma \\ -\widetilde{Y}_u & \widetilde{Y}_u^2 & \varrho\sigma\widetilde{Y}_u \\ -\varrho\sigma & \varrho\sigma\widetilde{Y}_u & \sigma^2 \end{bmatrix} \mathrm{d}u,$$

$$\widetilde{\boldsymbol{f}}_T := \frac{1}{(1-\varrho^2)\sigma^2} \begin{bmatrix} \int_0^T \frac{\mathrm{d}\widetilde{Y}_u}{\widetilde{Y}_u} - \varrho\sigma \int_0^T \frac{\mathrm{d}\widetilde{S}_u - \widetilde{S}_{u-}\,\mathrm{d}L_u}{\widetilde{Y}_u \widetilde{S}_{u-}} \\ -\int_0^T 1\,\mathrm{d}\widetilde{Y}_u + \varrho\sigma \int_0^T \frac{\mathrm{d}\widetilde{S}_u - \widetilde{S}_{u-}\,\mathrm{d}L_u}{\widetilde{S}_{u-}} \\ -\varrho\sigma \int_0^T \frac{\mathrm{d}\widetilde{Y}_u}{\widetilde{Y}_u} + \sigma^2 \int_0^T \frac{\mathrm{d}\widetilde{S}_u - \widetilde{S}_{u-}\,\mathrm{d}L_u}{\widetilde{Y}_u \widetilde{S}_{u-}} \end{bmatrix}.$$

Existence and uniqueness of MLE, III

One can see that for calculating $\widetilde{\boldsymbol{G}}_T$ and $\widetilde{\boldsymbol{f}}_T$, one does not need to know the parameters $\boldsymbol{\psi}$ and $\widetilde{\boldsymbol{\psi}}$, and hence the expression for the log-likelihood function obtained above can be extended to all $(\theta, \kappa, \mu) \in \mathbb{R}^3$ in a natural way.

Maximizing this extended function in $(\theta, \kappa, \mu) \in \mathbb{R}^3$, there exists a unique MLE $\widehat{\boldsymbol{\psi}}_T := (\widehat{\theta}_T, \widehat{\kappa}_T, \widehat{\mu}_T)$ of $\boldsymbol{\psi} = (\theta, \kappa, \mu)$ on $\mathbb{R}^3$ based on the observations $(S_t, L_t)_{t\in[0,T]}$ given by

$$\widehat{\theta}_T = \frac{\int_0^T Y_u \, \mathrm{d}u \int_0^T \frac{\mathrm{d}Y_u}{Y_u} - T\int_0^T \mathrm{d}Y_u + \varrho\sigma T \int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{S_{u-}} - \varrho\sigma T^2 \frac{\int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{Y_u S_{u-}}}{\int_0^T \frac{\mathrm{d}u}{Y_u}}}{T\int_0^T \frac{\mathrm{d}Y_u}{Y_u} - \int_0^T \frac{\mathrm{d}u}{Y_u}\int_0^T \mathrm{d}Y_u + \varrho\sigma \int_0^T \frac{\mathrm{d}u}{Y_u}\int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{S_{u-}} - \varrho\sigma T \int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{Y_u S_{u-}}},$$

$$\widehat{\kappa}_T = \frac{T\int_0^T \frac{\mathrm{d}Y_u}{Y_u} - \int_0^T \frac{\mathrm{d}u}{Y_u}\int_0^T \mathrm{d}Y_u + \varrho\sigma \int_0^T \frac{\mathrm{d}u}{Y_u}\int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{S_{u-}} - \varrho\sigma T \int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{Y_u S_{u-}}}{\int_0^T Y_u \, \mathrm{d}u \int_0^T \frac{\mathrm{d}u}{Y_u} - T^2},$$

$$\widehat{\mu}_T = \frac{\int_0^T \frac{\mathrm{d}S_u - S_{u-}\,\mathrm{d}L_u}{Y_u S_{u-}}}{\int_0^T \frac{\mathrm{d}u}{Y_u}}, \qquad \text{almost surely.}$$

Note that the distribution of $\widehat{\boldsymbol{\psi}}_T$ does not depend on $(L_t)_{t\geqslant 0}$, and $\widehat{\boldsymbol{\psi}}_T$ is a measurable function of $(S_t, L_t)_{t\in[0,T]}$.

Consistency of MLE

Theorem. If $\theta, \kappa > 0$ with $\theta\kappa > \frac{\sigma^2}{2}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\varrho \in (-1, 1)$, and $(Y_0, S_0) = (y_0, s_0) \in (0, \infty)^2$, then the MLE of $\boldsymbol{\psi} = (\theta, \kappa, \mu)$ is strongly consistent, i.e.,

$$\widehat{\boldsymbol{\psi}}_T = \big(\widehat{\theta}_T, \widehat{\kappa}_T, \widehat{\mu}_T\big) \xrightarrow{\text{a.s.}} \boldsymbol{\psi} = (\theta, \kappa, \mu) \qquad \text{as } T \to \infty.$$

If $\theta\kappa = \frac{\sigma^2}{2}$, then our forthcoming results for the asymptotic behaviour of the MLE of (θ, κ, μ) implies its weak consistency.

Asymptotic behaviour of MLE, I

Theorem. If $\theta, \kappa > 0$ with $\theta\kappa > \frac{\sigma^2}{2}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\varrho \in (-1, 1)$, and $(Y_0, S_0) = (y_0, s_0) \in (0, \infty)^2$, then the MLE of $\psi = (\theta, \kappa, \mu)$ is asymptotically normal, namely,

$$T^{1/2}(\widehat{\psi}_T - \psi) \xrightarrow{\mathcal{L}} \mathcal{N}_3(\mathbf{0}, \boldsymbol{V}) \qquad \text{as } T \to \infty,$$

where the matrix $\boldsymbol{V}$ is given by

$$\frac{1}{2\kappa^3}\begin{bmatrix} \sigma^2(2\theta\kappa - \varrho^2\sigma^2) & -2(1-\varrho^2)\sigma^2\kappa^2 & \varrho\sigma\kappa(2\theta\kappa - \sigma^2) \\ -2(1-\varrho^2)\sigma^2\kappa^2 & 4\kappa^4(1-\varrho^2) & 0 \\ \varrho\sigma\kappa(2\theta\kappa - \sigma^2) & 0 & \kappa^2(2\theta\kappa - \sigma^2) \end{bmatrix}.$$

Note that the limit covariance matrix $\boldsymbol{V}$ depends only on the unknown parameters θ and κ, but not on (the unknown) μ.

Asymptotic behaviour of MLE, II

Theorem. If $\theta, \kappa > 0$ with $\theta\kappa = \frac{\sigma^2}{2}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\varrho \in (-1, 1)$, and $(Y_0, S_0) = (y_0, s_0) \in (0, \infty)^2$, then

$$\begin{bmatrix} T^{1/2}(\widehat{\theta}_T - \theta) \\ T^{1/2}(\widehat{\kappa}_T - \kappa) \\ T(\widehat{\mu}_T - \mu) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} -\frac{\sigma^2\sqrt{1-\varrho^2}}{\sqrt{2\kappa^3}} Z_1 \\ \sqrt{2(1-\varrho^2)\kappa}\, Z_1 \\ \frac{\varrho\sigma}{\kappa\tau} + \frac{\sigma\sqrt{1-\varrho^2}}{\kappa\sqrt{\tau}} Z_2 \end{bmatrix} \qquad \text{as } T \to \infty,$$

where $\tau := \inf\{t \geqslant 0 : \mathcal{W}_t = 1\}$ with a standard Wiener process $(\mathcal{W}_t)_{t\geqslant 0}$, and Z_1 and Z_2 are independent standard normally distributed random variables, independent from τ.

The limit distribution above is mixed normal.

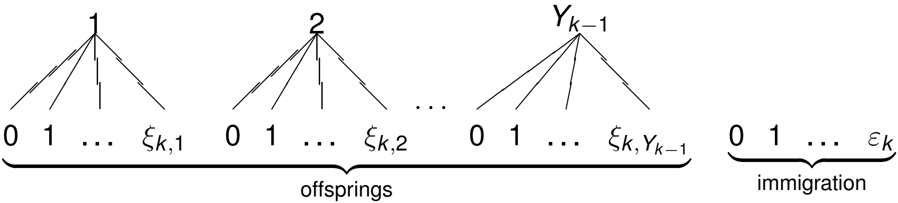
This talk is based on:

BARCZY, M. PAP, G.,
Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations.
Statistics **50(2)** (2016), 389–417.

BARCZY, M., BEN ALAYA, M., KEBAIER, A., PAP, G.,
Asymptotic behavior of maximum likelihood estimators for a jump-type Heston model. Submitted.
Arxiv: 1509.08869 (2016).

Thank you for your attention!

Galton-Watson branching processes with immigration



$$Y_k = \sum_{j=1}^{Y_{k-1}} \xi_{k,j} + \varepsilon_k, \quad k \in \mathbb{N} := \{1, 2, \ldots\}, \quad Y_0 = \eta_0,$$

where

- $\{\eta_0, \xi_{k,j}, \varepsilon_k : k, j \in \mathbb{N}\}$ are independent rv's with values in $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$,
- $\{\xi_{k,j} : k, j \in \mathbb{N}\}$ are identically distributed,
- $\{\varepsilon_k : k \in \mathbb{N}\}$ are identically distributed.

Scaling limits of GWI processes: CBI processes

Theorem. (Kawazu and Watanabe (1971), Li (2006)) For all $n \in \mathbb{N}$, let $(\zeta_k^{(n)})_{k\in\mathbb{Z}_+}$ be a GWI process, and $\gamma_n > 0$ with $\gamma_n \uparrow \infty$. Under some conditions on the generator functions of the offspring and immigration distributions,

$$(n^{-1}\zeta^{(n)}_{\lfloor \gamma_n t\rfloor})_{t\geqslant 0} \stackrel{\mathcal{L}}{\longrightarrow} (Y_t)_{t\geqslant 0} \qquad \text{as } n \to \infty,$$

where the limit process Y is a CBI process, i.e., it is a time-homogeneous Markov process with state space $[0, \infty)$ and (conditional) Laplace transform

$$\mathrm{E}\left(\mathrm{e}^{-\lambda Y_t} \mid Y_0 = y\right) = \exp\left\{-v_t(\lambda)y - \int_0^t \psi(v_s(\lambda))\,\mathrm{d}s\right\}$$

for $\lambda, y, t \geqslant 0$, where $(v_t(\lambda))_{t\geqslant 0}$ and ψ are uniquely determined non-negative functions.

For any $t \geqslant 0$, the (cond.) Laplace transform of Y_t depends exponentially affine on the initial value y (i.e., of the form $\exp\{c_1 y + c_2\}$).

Another two-factor affine diffusion model

Let $a > 0$, $b, m, \theta \in \mathbb{R}$, and let us consider the SDE

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (m - \theta X_t)\,\mathrm{d}t + \sqrt{Y_t}\,\mathrm{d}B_t, \end{cases} \qquad t \geqslant 0,$$

where $(W_t)_{t\geqslant 0}$ and $(B_t)_{t\geqslant 0}$ are independent standard Wiener processes.

Chen and Joslin (2012): applications of this model in financial mathematics.

On the observation X for MLE

We explain why we suppose only that the process X is observed for MLE, and not (Y, X).

If $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then, by the SDE for (Y, X),

$$\langle X \rangle_t = \sigma_2^2 \int_0^t Y_s \, \mathrm{d}s, \qquad t \geqslant 0.$$

It is known that,

$$\sum_{i=1}^{\lfloor nt \rfloor} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2 \xrightarrow{\mathrm{P}} \langle X \rangle_t \qquad \text{as } n \to \infty, \quad t \geqslant 0.$$

This convergence holds almost surely along a suitable subsequence, the members of this sequence are measurable functions of $(X_s)_{s \in [0,t]}$, hence $\langle X \rangle_t = \sigma_2^2 \int_0^t Y_s \, \mathrm{d}s$ is a measurable function of $(X_s)_{s \in [0,t]}$.

Moreover,

$$\frac{\langle X\rangle_{t+h} - \langle X\rangle_t}{h} = \frac{\sigma_2^2}{h}\int_t^{t+h} Y_s\,\mathrm{d}s \xrightarrow{\text{a.s.}} \sigma_2^2 Y_t \qquad \text{as } h \to 0, \quad t \geqslant 0,$$

since Y has almost surely continuous sample paths. In particular,

$$\frac{\langle X\rangle_h}{h y_0} = \frac{\sigma_2^2}{h y_0}\int_0^h Y_s\,\mathrm{d}s \xrightarrow{\text{a.s.}} \sigma_2^2 \frac{Y_0}{y_0} = \sigma_2^2 \qquad \text{as } h \to 0,$$

hence, for any fixed $T > 0$, σ_2^2 is a measurable function of $(X_s)_{s\in[0,T]}$, i.e., it can be determined from a sample $(X_s)_{s\in[0,T]}$.

Consequently, for all $t \in [0, T]$, Y_t is a measurable function of $(X_s)_{s\in[0,T]}$, i.e., it can be determined from a sample $(X_s)_{s\in[0,T]}$.

Next we give statistics for the parameters σ_1, σ_2 and ϱ using continuous time observations $(X_t)_{t\in[0,T]}$ with some $T > 0$. Due to this result we do not consider the estimation of these parameters, they are supposed to be known.

Statistics for σ_1, σ_2 and ϱ

If $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then for all $T > 0$,

$$\boldsymbol{S} := \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \frac{1}{\int_0^T Y_s \, \mathrm{d}s} \begin{bmatrix} \langle Y \rangle_T & \langle Y, X \rangle_T \\ \langle Y, X \rangle_T & \langle X \rangle_T \end{bmatrix} =: \widehat{\boldsymbol{S}}_T \quad \text{a.s.},$$

where $(\langle Y, X \rangle_t)_{t \geqslant 0}$ is the quadratic co-variation of Y and X.

Here $\widehat{\boldsymbol{S}}_T$ is a statistic, i.e., $\exists$ a measurable function $\Xi : C([0, T], [0, \infty) \times \mathbb{R}) \to \mathbb{R}^{2 \times 2}$ with $\widehat{\boldsymbol{S}}_T = \Xi((X_s)_{s \in [0, T]})$.

Indeed, as $n \to \infty$,

$$\frac{1}{\frac{1}{n} \sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}}} \sum_{i=1}^{\lfloor nT \rfloor} \begin{bmatrix} Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \\ X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \end{bmatrix} \begin{bmatrix} Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \\ X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \end{bmatrix}^\top \xrightarrow{\mathrm{P}} \widehat{\boldsymbol{S}}_T.$$

Existence and uniqueness of solution, affine process

- $(Y_t, X_t)_{t \geqslant 0}$ is an affine process with infinitesimal generator

$$
\begin{aligned}
(\mathcal{A}f)(y,x) &= (a - by) f_1'(y,x) + (\alpha - \beta y) f_2'(y,x) \\
&\quad + \frac{1}{2} y \big(\sigma_1^2 f_{1,1}''(y,x) + 2\varrho \sigma_1 \sigma_2 f_{1,2}''(y,x) + \sigma_2^2 f_{2,2}''(y,x) \big),
\end{aligned}
$$

where $(y,x) \in [0,\infty) \times \mathbb{R}$ and $f \in C_c^2([0,\infty) \times \mathbb{R})$.

The infinitesimal generator of (Y, X) is

$$(\mathcal{A}f)(y,x) := \lim_{t \downarrow 0} \frac{\mathrm{E}(f(Y_t, X_t) \mid (Y_0, X_0) = (y,x)) - f(y,x)}{t}, \qquad y \geqslant 0,\ x \in \mathbb{R}$$

for suitable functions $f : [0,\infty) \times \mathbb{R} \to \mathbb{R}$.

We note that

- if $a \geqslant \sigma_1^2/2$, then $\mathrm{P}(Y_t > 0 \text{ for all } t > 0) = 1$.
- if $0 < a < \sigma_1^2/2$, then Y hits 0 with probability $p \in (0, 1)$ in case of $b < 0$, and 1 in case of $b \geqslant 0$ (and zero is reflecting).

Stationarity and ergodicity in the subcritical case, I

Theorem. (Feller (1951); Cox, Ingersoll and Ross (1985)) Let $a, b, \sigma_1 > 0$. Let $(Y_t)_{t \geqslant 0}$ be the unique strong solution of the first equation of the given SDE satisfying $\mathrm{P}(Y_0 \geqslant 0) = 1$.

(i) Then $Y_t \xrightarrow{\mathcal{L}} Y_\infty$ as $t \to \infty$, and the distribution of Y_∞ is given by

$$\mathrm{E}(\mathrm{e}^{-\lambda Y_\infty}) = \left(1 + \frac{\sigma_1^2}{2b}\lambda\right)^{-2a/\sigma_1^2}, \qquad \lambda \geqslant 0,$$

i.e., Y_∞ has Gamma distribution with parameters $2a/\sigma_1^2$ and $2b/\sigma_1^2$.

Especially, if $a \in \left(\frac{\sigma_1^2}{2}, \infty\right)$, then $\mathrm{E}\left(\frac{1}{Y_\infty}\right) = \frac{2b}{2a - \sigma_1^2}$.

Stationarity and ergodicity in the subcritical case, II

(ii) If the random initial value Y_0 has the same distribution as Y_∞, then the process $(Y_t)_{t\geqslant 0}$ is strictly stationary.

(iii) For all Borel measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that $\mathrm{E}(|f(Y_\infty)|) < \infty$, we have

$$\frac{1}{T}\int_0^T f(Y_s)\,\mathrm{d}s \xrightarrow{\text{a.s.}} \mathrm{E}(f(Y_\infty)) \qquad \text{as } T \to \infty,$$

and

$$\frac{1}{n}\sum_{i=0}^{n-1} f(Y_i) \xrightarrow{\text{a.s.}} \mathrm{E}(f(Y_\infty)) \qquad \text{as } n \to \infty.$$

From now on, we consider the Heston model with a non-random initial value $(y_0, x_0) \in (0,\infty) \times \mathbb{R}$, and we equip $(\Omega, \mathcal{F}, \mathrm{P})$ with the augmented filtration $(\mathcal{F}_t)_{t\geqslant 0}$ corresponding to $(W_t, B_t)_{t\geqslant 0}$, constructed as in Karatzas and Shreve (1991).

Consistency of MLE: supercritical case $(b < 0)$

Theorem. If $a \in \left[\frac{\sigma_1^2}{2}, \infty\right)$, $b < 0$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(Y_0, X_0) = (y_0, x_0) \in (0, \infty) \times \mathbb{R}$, then the MLE of b is strongly consistent, i.e., $\widehat{b}_T \xrightarrow{\text{a.s.}} b$ as $T \to \infty$.

The proof is based on

- $e^{bt} Y_t$ converges almost surely as $t \to \infty$.
- an integral version of the Kronecker Lemma due to Küchler and Sørensen (1997): Let $a : [0, \infty) \to [0, \infty)$ be a measurable function. Suppose that $\lim_{T\to\infty} \int_0^T a(t)\,\mathrm{d}t = \infty$. Then for every bounded and measurable function $f : [0, \infty) \to \mathbb{R}$ for which the limit $\lim_{t\to\infty} f(t) =: f(\infty)$ exists, we have

$$\lim_{T\to\infty} \frac{\int_0^T a(t) f(t)\,\mathrm{d}t}{\int_0^T a(t)\,\mathrm{d}t} = f(\infty).$$

It will turn out that for supercritical Heston models with $a \in \left[\sigma_1^2/2, \infty\right)$, the MLE of a and α is not even weakly consistent, but the MLE of β is weakly consistent.

Asymptotics of MLE: subcritical, diffusion case $(b > 0)$

With a random scaling,

$$\frac{1}{\left(\int_0^T \frac{\mathrm{d}s}{Y_s}\right)^{1/2}} \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} \int_0^T \frac{\mathrm{d}s}{Y_s} & -T \\ 0 & \left(\int_0^T Y_s \, \mathrm{d}s \int_0^T \frac{\mathrm{d}s}{Y_s} - T^2\right)^{1/2} \end{bmatrix} \right) \begin{bmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \\ \widehat{\alpha}_T - \alpha \\ \widehat{\beta}_T - \beta \end{bmatrix}$$

$$\xrightarrow{\mathcal{L}} \mathcal{N}_4 \left(\boldsymbol{0}, \boldsymbol{S} \otimes \boldsymbol{I}_2\right) \qquad \text{as} \ \ T \to \infty,$$

where $\boldsymbol{I}_2$ is the 2×2 unit matrix.

Asymptotics of MLE: critical, diffusion case $(b = 0)$

With a random scaling,

$$\begin{bmatrix} \left(\int_0^T \frac{\mathrm{d}s}{Y_s}\right)^{1/2} (\widehat{a}_T - a) \\ \left(\int_0^T \frac{\mathrm{d}s}{Y_s}\right)^{1/2} (\widehat{\alpha}_T - \alpha) \\ \left(\int_0^T Y_s \,\mathrm{d}s\right)^{1/2} \widehat{b}_T \\ \left(\int_0^T Y_s \,\mathrm{d}s\right)^{1/2} (\widehat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \boldsymbol{S}^{1/2} \boldsymbol{Z}_2 \\ \frac{a - \mathcal{Y}_1}{\left(\int_0^1 \mathcal{Y}_s \,\mathrm{d}s\right)^{1/2}} \\ \frac{\alpha - \mathcal{X}_1}{\left(\int_0^1 \mathcal{Y}_s \,\mathrm{d}s\right)^{1/2}} \end{bmatrix} \qquad \text{as } T \to \infty.$$

Asymptotics of MLE: supercritical, diffusion case ($b < 0$)

With a random scaling,

$$\begin{bmatrix} \widehat{a}_T - a \\ \widehat{\alpha}_T - \alpha \\ \left(\int_0^T Y_s \, ds\right)^{1/2} (\widehat{b}_T - b) \\ \left(\int_0^T Y_s \, ds\right)^{1/2} (\widehat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \widetilde{\mathcal{V}} \\ \varrho \frac{\sigma_2}{\sigma_1} \widetilde{\mathcal{V}} + \sigma_2 \sqrt{1 - \varrho^2} \left(\int_0^{-1/b} \widetilde{\mathcal{Y}}_u \, du\right)^{-1/2} Z_1 \\ \boldsymbol{S}^{1/2} \boldsymbol{Z}_2 \end{bmatrix}$$

as $T \to \infty$.

Subcritical case $(b > 0)$: asymptotics of MLE, sketch of proof

It is based on the decomposition:

$$\sqrt{T}\begin{bmatrix}\widehat{a}_T - a\\ \widehat{b}_T - b\\ \widehat{\alpha}_T - \alpha\\ \widehat{\beta}_T - \beta\end{bmatrix} = \left(\boldsymbol{I}_2 \otimes \begin{bmatrix}\frac{1}{T}\int_0^T \frac{\mathrm{d}s}{Y_s} & -1\\ -1 & \frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\end{bmatrix}^{-1}\right)\frac{1}{\sqrt{T}}\boldsymbol{M}_T,$$

where

$$\boldsymbol{M}_t := \left[\sigma_1\int_0^t \frac{\mathrm{d}W_s}{\sqrt{Y_s}},\quad -\sigma_1\int_0^t \sqrt{Y_s}\,\mathrm{d}W_s,\quad \sigma_2\int_0^t \frac{\mathrm{d}\widetilde{W}_s}{\sqrt{Y_s}},\quad -\sigma_2\int_0^t \sqrt{Y_s}\,\mathrm{d}\widetilde{W}_s\right]^\top$$

is a continuous local martingale. The first term of the decomposition above converges a.s. to a deterministic matrix, while the second term in distribution to a 4-dimensional normally distributed random variable due to the martingale central limit theorem, since $\frac{1}{T}\langle M\rangle_T$ converges a.s. as $T \to \infty$.

Critical case $(b = 0)$: asymptotics of MLE, sketch of proof

Consider the difference of the MLE and the true parameter vector, an *appropriate* decomposition of it, the joint characteristic function of the terms in this decomposition and study its asymptotic behavior as $T \to \infty$ using that:

- Overbeck (1998):

$$\frac{1}{\log T} \int_0^T \frac{\mathrm{d}s}{Y_s} \xrightarrow{\mathrm{P}} \left(a - \frac{\sigma_1^2}{2}\right)^{-1} \qquad \text{as } T \to \infty,$$

- Ben Alaya and Kebaier (2013):

$$\left(\frac{\log Y_T - \log y_0 + \left(\frac{\sigma_1^2}{2} - a\right) \int_0^T \frac{\mathrm{d}s}{Y_s}}{\sqrt{\log T}}, \frac{Y_T}{T}, \frac{1}{T^2} \int_0^T Y_s \, \mathrm{d}s \right)$$

$$\xrightarrow{\mathcal{L}} \left(\frac{\sigma_1}{\sqrt{a - \frac{\sigma_1^2}{2}}} Z_1, \mathcal{Y}_1, \int_0^1 \mathcal{Y}_s \, \mathrm{d}s \right) \qquad \text{as } T \to \infty,$$

where $Z_1 = (\boldsymbol{Z}_2)_{1,1}$ is a 1-dimensional standard normally dist. r.v. independent of $(\mathcal{Y}_1, \int_0^1 \mathcal{Y}_t \, \mathrm{d}t)$.

Supercritical case $(b < 0)$: asymptotics of MLE, sketch of proof

Consider the difference of the MLE and the true parameter vector, an *appropriate* decomposition of it, the joint characteristic function of the terms in this decomposition and study its asymptotic behavior as $T \to \infty$ using that:

$$\left(\frac{\int_0^T \sqrt{Y_s}\,\mathrm{d}W_s}{\left(\int_0^T Y_s\,\mathrm{d}s\right)^{1/2}}, \mathrm{e}^{bT} Y_T, \mathrm{e}^{bT} \int_0^T Y_s\,\mathrm{d}s, \int_0^T \frac{\mathrm{d}s}{Y_s} \right)$$

$$\xrightarrow{\mathcal{L}} \left(Z_2, \widetilde{\mathcal{Y}}_{-1/b}, -\frac{\widetilde{\mathcal{Y}}_{-1/b}}{b}, \int_0^{-1/b} \widetilde{\mathcal{Y}}_u\,\mathrm{d}u \right) \qquad \text{as } T \to \infty,$$

where $Z_2 = (\boldsymbol{Z}_2)_{1,1}$ is a 1-dimensional standard normally distributed r.v. independent of $\left(\widetilde{\mathcal{Y}}_{-1/b}, \int_0^{-1/b} \widetilde{\mathcal{Y}}_u\,\mathrm{d}u\right)$.

Stationarity and ergodicity, original parametrization

Theorem. (Feller (1951); Cox, Ingersoll and Ross (1985)) Let $\theta, \kappa, \sigma > 0$. Let $(Y_t)_{t \geqslant 0}$ be the unique strong solution of the first equation of the given SDE satisfying $\mathsf{P}(Y_0 \geqslant 0) = 1$.

- Then $Y_t \xrightarrow{\mathcal{L}} Y_\infty$ as $t \to \infty$, and the distribution of Y_∞ is given by

$$\mathsf{E}(\mathrm{e}^{-\lambda Y_\infty}) = \left(1 + \frac{\sigma^2}{2\kappa}\lambda\right)^{-\frac{2\theta\kappa}{\sigma^2}}, \qquad \lambda \geqslant 0,$$

  i.e., Y_∞ has Gamma distribution with parameters $2\theta\kappa/\sigma^2$ and $2\kappa/\sigma^2$, hence

$$\mathsf{E}(Y_\infty^K) = \frac{\Gamma\left(\frac{2\theta\kappa}{\sigma^2} + K\right)}{\left(\frac{2\kappa}{\sigma^2}\right)^K \Gamma\left(\frac{2\theta\kappa}{\sigma^2}\right)}, \qquad K \geqslant -\frac{2\theta\kappa}{\sigma^2}.$$

  Further, if $\theta\kappa \geqslant \frac{\sigma^2}{2}$, then $\mathsf{E}\left(\frac{1}{Y_\infty}\right) = \frac{2\kappa}{2\theta\kappa - \sigma^2}$.
- For all Borel measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that $\mathsf{E}(|f(Y_\infty)|) < \infty$, we have

$$\frac{1}{T}\int_0^T f(Y_u)\,\mathrm{d}u \xrightarrow{\text{a.s.}} \mathsf{E}(f(Y_\infty)) \qquad \text{as } T \to \infty.$$

On the sample and on the parameters σ and ϱ

We do not suppose that the process $(Y_t)_{t\in[0,T]}$ is observed, since, for all $t \in [0, T]$, Y_t is a measurable function (i.e., a statistic) of $(S_t)_{t\in[0,T]}$.

Due to similar reasons, we do not estimate σ and ϱ.

If

$$L_t = \sum_{s\in[0,t]} \Delta L_s, \qquad t \geqslant 0,$$

then for all $t \in [0, T]$, L_t is a measurable function of $(S_u)_{u\in[0,T]}$, so (theoretically) we do not have to suppose that L is observed.

A sufficient condition for this:
$\int_{-1}^{1} |z|\, m(\mathrm{d}z) < \infty$ and $\gamma = \int_{-1}^{1} z\, m(\mathrm{d}z)$.

It turns out that for calculating the MLE of (θ, κ, μ), one does not need to know the parameter γ and the Lévy measure m.

Asymptotic behaviour of MLE, jump-type case, $\theta\kappa > \frac{\sigma^2}{2}$

With a random scaling, we have

$$\boldsymbol{R}_T \boldsymbol{Q}_T(\widehat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}) \xrightarrow{\mathcal{L}} \mathcal{N}_3(\boldsymbol{0}, \boldsymbol{I}_3) \qquad \text{as} \quad T \to \infty,$$

where $\boldsymbol{I}_3$ denotes the 3×3 identity matrix, and $\boldsymbol{R}_T$, $T > 0$, and $\boldsymbol{Q}_T$, $T > 0$, are 3×3 (not uniquely determined) random matrices with properties $T^{-1/2}\boldsymbol{R}_T \xrightarrow{\mathrm{P}} \boldsymbol{C}$ as $T \to \infty$ with some $\boldsymbol{C} \in \mathbb{R}^{3\times 3}$, $\boldsymbol{R}_T^\top \boldsymbol{R}_T = \boldsymbol{G}_T$, $T > 0$, and $\boldsymbol{Q}_T \xrightarrow{\mathrm{P}} \boldsymbol{Q}$ as $T \to \infty$, where

$$\boldsymbol{Q} := \begin{bmatrix} \kappa & \theta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{G}_T := \frac{1}{(1-\varrho^2)\sigma^2} \int_0^T \frac{1}{Y_u} \begin{bmatrix} 1 & -Y_u & -\varrho\sigma \\ -Y_u & Y_u^2 & \varrho\sigma Y_u \\ -\varrho\sigma & \varrho\sigma Y_u & \sigma^2 \end{bmatrix} \mathrm{d}u.$$

One can choose $\boldsymbol{R}_T$ and $\boldsymbol{Q}_T$ to be upper triangular matrices (in case of $\boldsymbol{R}_T$, it is the Cholesky factorization).

The advantage of the random scaling is that the limit covariance matrix above is $\boldsymbol{I}_3$, which does not depend on any of the unknown parameters.

Asymptotic behaviour of MLE, jump-type case, $\theta\kappa = \frac{\sigma^2}{2}$

With a random scaling, we have

$$\begin{bmatrix} \frac{\sigma T^2}{2\sqrt{1-\varrho^2}\left(\int_0^T Y_u \, \mathrm{d}u\right)^{3/2}} (\widehat{\theta}_T - \theta) \\ \frac{1}{\sigma\sqrt{1-\varrho^2}} \left(\int_0^T Y_u \, \mathrm{d}u\right)^{1/2} (\widehat{\kappa}_T - \kappa) \\ \frac{\sigma T^2}{2\int_0^T Y_u \, \mathrm{d}u} (\widehat{\mu}_T - \mu) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} -Z_1 \\ Z_1 \\ \frac{\varrho}{\tau} + \frac{\sqrt{1-\varrho^2}}{\sqrt{\tau}} Z_2 \end{bmatrix} \quad \text{as} \quad T \to \infty.$$

References

BEN ALAYA, M. AND KEBAIER, A.,
Parameter estimation for the square root diffusions: ergodic and nonergodic cases.
Stochastic Models **28(4)** (2012), 609–634.

BEN ALAYA, M. AND KEBAIER, A.,
Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions.
Stochastic Analysis and Applications **31(4)** (2013), 552–573

COX, J. C., INGERSOLL, J. E. AND ROSS, S. A.,
A theory of the term structure of interest rates.
Econometrica **53(2)** (1985), 385–407.

DAWSON, D. A. AND LI, Z.,
Skew convolution semigroups and affine Markov processes.
The Annals of Probability **34** (2006), 1103–1142.

References

Duffie, D., Filipović, D. and Schachermayer, W.,
Affine processes and applications in finance.
Annals of Applied Probability **13** (2003), 984–1053.

Heston, S.,
A closed-form solution for options with stochastic volatilities with applications to bond and currency options.
The Review of Financial Studies **6** (1993), 327–343.

Overbeck, L.,
Estimation for continuous branching processes.
Scandinavian Journal of Statistics **25(1)** (1998), 111–126.