Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model

Joon Y. Park

Department of Economics

Indiana University

Joint with Minsoo Jeong

Contents

Basics and Background

Maximum Likelihood Estimation

Continuous Time Asymptotics
- Preliminaries
- Limit Theories

Asymptotic Theory of MLE
- Primary Asymptotics
- Limit Distributions

Simulation Results
- Finite Sample Distributions
- Bias and Size Corrections

Basics and Background

Objective

Our objective is to develop the general asymptotic theory for maximum likelihood estimation, which is

- Not restricted to specific models and estimators
- Applicable for nonstationary, as well as stationary, models
- Useful for approximating finite sample distributions,

in contrast to the existing theory.

Two-Dimensional Asymptotics

Our asymptotics are obtained as $\Delta \to 0$ and $T \to \infty$, which has some clear advantages. It provides

- Unifying framework for stationary and nonstationary models
- Important contrast between the asymptotics of drift and diffusion parameters
- Theory applicable for a variety of MLE's relying on approximate transitions, which require $\Delta \to 0$

More importantly, they yield primary asymptotics that are practically more relevant and useful. Under stationarity, our asymptotics become identical to the conventional one-dimensional asymptotics.

Diffusion

The parametric diffusion model is defined as a solution to the stochastic differential equation (SDE)

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

where α and β are unknown parameters, W is standard Brownian motion, and

- $\mu(x, \alpha)$: *drift function* specifying instantaneous mean
- $\sigma(x, \beta)$: *diffusion function* specifying instantaneous volatility

of dX_t at $X_t = x$. Let (X_t) take values on $\mathcal{D} = (\underline{x}, \overline{x})$, where $\mathcal{D} = (-\infty, \infty)$ or $(0, \infty)$ in most cases.

Recurrence

For a diffusion X, define the hitting time of a point $y \in \mathcal{D}$ as

$$\tau_y = \inf\{t \geq 0 | X_t = y\}.$$

We say that a diffusion is *recurrent* if

$$\mathbb{P}\{\tau_y < \infty | X_0 = x\} = 1$$

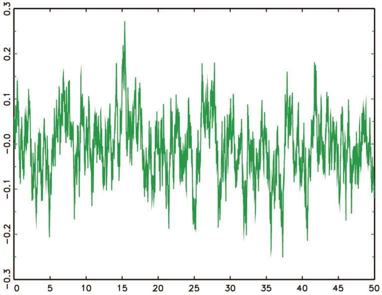
for all x and y in the interior of $\mathcal{D}$. A recurrent diffusion is said to be *positive recurrent* if $\mathbb{E}\{\tau_y < \infty | X_0 = x\} < \infty$, and *null recurrent* if $\mathbb{E}\{\tau_y < \infty | X_0 = x\} = \infty$. A diffusion which is not recurrent is said to be *transient*.

Recurrence and Stationarity

$$\text{Diffusion} \left\{ \begin{array}{l} \text{Positive recurrent} \\ \text{Null recurrent} \\ \text{Transient} \end{array} \right. \quad \begin{array}{l} \longrightarrow \text{Stationary} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \longrightarrow \text{Nonstationary} \end{array}$$

Recurrence and Stationarity

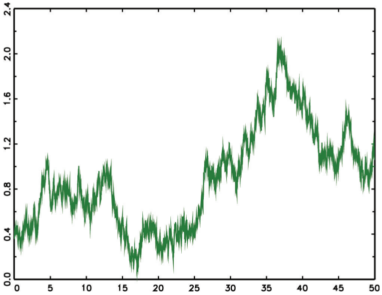
$$\text{Diffusion} \left\{ \begin{array}{l} \text{Positive recurrent} \\ \text{Null recurrent} \\ \text{Transient} \end{array} \right. \quad \begin{array}{l} \longrightarrow \text{Stationary} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \longrightarrow \text{Nonstationary} \end{array}$$



Positive recurrent process

Recurrence and Stationarity

$$\text{Diffusion} \begin{cases} \text{Positive recurrent} & \longrightarrow \text{Stationary} \\ \left.\begin{array}{l} \text{Null recurrent} \\ \text{Transient} \end{array}\right\} & \longrightarrow \text{Nonstationary} \end{cases}$$

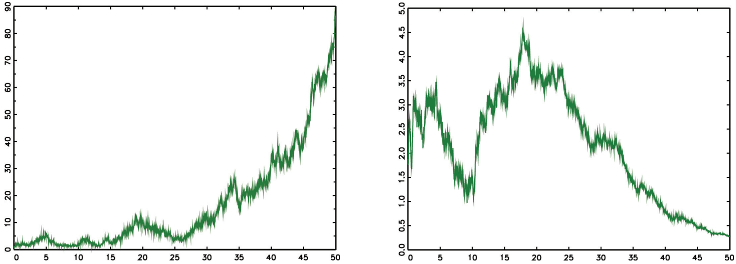


Null recurrent process

Recurrence and Stationarity

Diffusion $\begin{cases} \text{Positive recurrent} & \to \text{Stationary} \\ \text{Null recurrent} & \\ \text{Transient} & \end{cases}$

Null recurrent, Transient $\Big\} \to$ Nonstationary

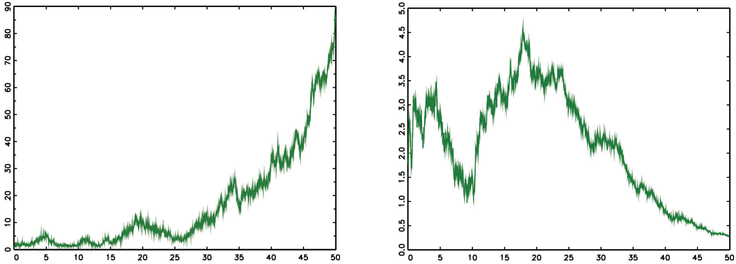


Transient process

Recurrence and Stationarity

Diffusion $\left\{ \begin{array}{l} \text{Positive recurrent} \\ \text{Null recurrent} \\ \text{Transient} \end{array} \right.$

- Positive recurrent $\longrightarrow$ Stationary
- Null recurrent, Transient $\Big\} \longrightarrow$ Nonstationary



Transient process

Examples

Some commonly used diffusion models are

- Brownian Motion with Drift: $dX_t = \alpha dt + \beta dW_t$
- Geometric Brownian Motion: $dX_t = \alpha X_t dt + \beta X_t dW_t$
- Ornstein-Uhlenbeck Process: $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta dW_t$
- Feller's Square-Root Process: $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta\sqrt{X_t}dW_t$
- Constant Elasticity of Variance Process:
  $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta_1 X_t^{\beta_2} dW_t$
- Nonlinear Drift Diffusion Process:
  $dX_t = (\alpha_1 + \alpha_2 X_t + \alpha_3 X_t^2 + \alpha_4 X_t^{-1})dt + \sqrt{\beta_1 + \beta_2 X_t + \beta_3 X_t^{\beta_4}} dW_t$

whose recurrence and stationarity properties are dependent upon parameter values.

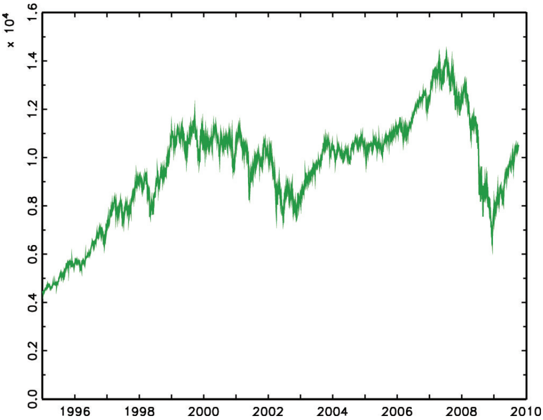
Financial Time Series

Plots of three major financial time series given by

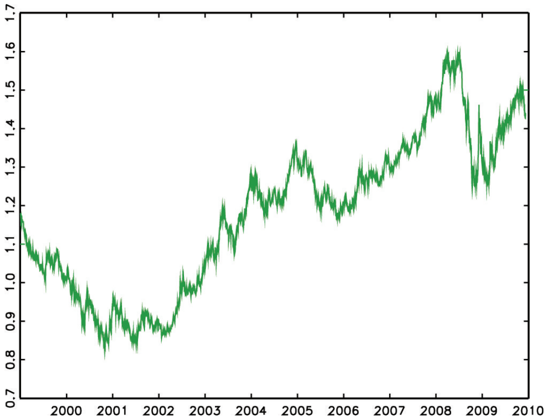
- Stock Index: Dow-Jones Industrial Average
- Exchange Rate: Euro-Dollar Exchange Rate
- Interest Rate: Federal Funds Rate

Many other financial time series show very similar patterns and characteristics.

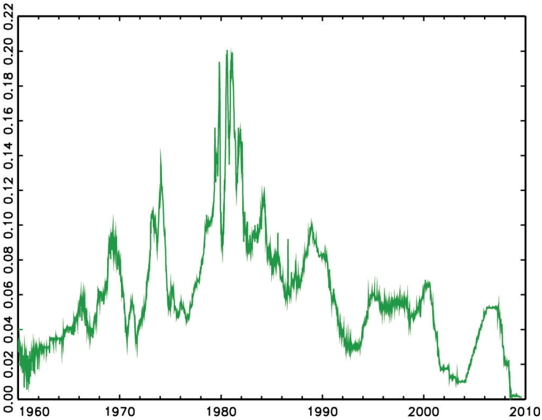
Dow-Jones Industrial Average Index



Euro-Dollar Exchange Rates



Federal Funds Rates



Empirical Facts

We may summarize them as

- All important financial time series are either nonstationary or near-nonstationary
- Existing nonstationarity cannot be fully removed by simple detrending or transformations
- Null recurrent processes are of most practical relevancy

Some of financial time series appear to be transient. However, in many cases we may readily reduce them to nonstationary recurrent processes by simple detrending or transformations.

Scale Function

The scale function is defined as

$$s(x) = \int_w^x \exp\left(- \int_w^y \frac{2\mu(z)}{\sigma^2(z)}\, dz \right) dy$$

for some $w \in \mathcal{D}$, which is a solution of the ODE $\mu \dot{s} + \sigma^2 \ddot{s}/2 = 0$. It follows from Ito formula that

$$ds(X_t) = \left(\mu \dot{s} + \sigma^2 \ddot{s}/2\right)(X_t)dt + \left(\sigma \dot{s}\right)(X_t)dW_t = \left(\sigma \dot{s}\right)(X_t)dW_t,$$

and (X_t^s), $X_t^s = s(X_t)$, becomes a martingale, which is said to be in natural scale. The scale function is defined uniquely up to any affine transformations.

Recurrence

A diffusion (X_t) on $\mathcal{D} = (-\underline{x}, \overline{x})$ is *recurrent* if and only if

$$s(\underline{x}) = -\infty \quad \text{and} \quad s(\overline{x}) = \infty$$

under the usual regularity conditions that are satisfied for most commonly used diffusion models. Note that the scale function is strictly increasing. A diffusion in natural scale has the identity scale function.

Speed Measure and Stationarity

The speed measure is given by the density

$$m(x,\theta) = \frac{2}{(\sigma^2 \dot{s})(x,\theta)}$$

with respect to the Lebesgue measure. A diffusion is positive recurrent if and only if $m(\mathcal{D},\theta) < \infty$, and null recurrent if and only if $m(\mathcal{D},\theta) = \infty$. For a positive recurrent diffusion,

$$\pi(x,\theta) = \frac{m(x,\theta)}{m(\mathcal{D},\theta)}$$

becomes the time invariant stationary density.

Maximum Likelihood Estimation

Maximum Likelihood Estimator

For diffusion process (X_t), the samples of size n are assumed to be collected at interval Δ over time horizon T, i.e.,

$$x_\Delta, x_{2\Delta}, \ldots, x_{n\Delta}$$

with $T = n\Delta$. The maximum likelihood estimator $\hat{\theta}$ of θ is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log p(\Delta, x_{(i-1)\Delta}, x_{i\Delta}, \theta)$$

where p is the transition density. In most cases, the exact transition density is unknown and has to be approximated or simulated.

Euler Approximation

The diffusion model is often approximated by

$$X_{i\Delta} - X_{(i-1)\Delta} \simeq \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}),$$

which yields the transition

$$\mathcal{L}(X_{i\Delta}|X_{(i-1)\Delta} = x) = \mathbb{N}\left(x + \Delta\mu(x), \Delta\sigma^2(x)\right).$$

The transition based on the Euler approximation is therefore normal.

Milstein Approximation

A better approximation may be obtained from

$$X_{i\Delta} - X_{(i-1)\Delta} \simeq \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + \frac{1}{2}\sigma\dot{\sigma}(X_{(i-1)\Delta})\left[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta\right],$$

which yields the transition

$$\mathcal{L}(X_{i\Delta}|X_{(i-1)\Delta} = x) = \mathcal{L}\left(x + \Delta\mu(x) + \sqrt{\Delta}\sigma(x)\mathbb{N}(0,1) + \frac{\Delta}{2}\sigma\dot{\sigma}(x)\left[\mathbb{N}(0,1)^2 - 1\right]\right),$$

where $\mathbb{N}(0,1)$ is a standard normal random variate.

Aït-Sahalia (2002) Approach

First, define (X_t^*) from (X_t) using the Lamperti transformation given by $\tau(x,\beta) = \int^x dy/\sigma(y,\beta)$, so that we have

$$dX_t^* = \left[\frac{\mu\left(\tau^{-1}(X_t^*,\beta),\alpha\right)}{\sigma\left(\tau^{-1}(X_t^*,\beta),\beta\right)} - \frac{1}{2}\dot{\sigma}\left(\tau^{-1}(X_t^*,\beta),\beta\right) \right] dt + dW_t,$$

whose transition density is denoted by p^*.

Second, consider the transition

$$\mathcal{L}\left(X_{i\Delta}^{**} = \Delta^{-1/2}(X_{i\Delta}^* - X_{(i-1)\Delta}^*) \middle| X_{(i-1)\Delta}^* = x \right)$$

and write its density as p^{**}.

Third, approximate p^{**} using the Hermite expansion

$$p^{**}(\Delta, x, y, \theta) \simeq p_K^{**}(\Delta, x, y, \theta) = \phi(y) \sum_{k=0}^{K} c_k(\Delta, x, \theta) H_k(y),$$

where (H_k) are Hermite polynomials and (c_k) are coefficients obtained from the approximated conditional moments of (X_t^*).

The original transition density can now be approximated by

$$p_{AS}(\Delta, x, y, \theta) = \frac{1}{\sqrt{\Delta}\sigma(y, \beta)} p_K^{**} \left(\Delta, \tau(x, \beta), \frac{\tau(y, \beta) - \tau(x, \beta)}{\sqrt{\Delta}}, \theta \right),$$

which we may use to do the maximum likelihood estimation.

Quasi-ML Estimation

It is also possible to use the quasi-maximum likelihood approach based on the approximated transition mean and variance. If the conditional mean and variance of $X_{i\Delta} - X_{(i-1)\Delta}$ given $X_{(i-1)\Delta} = x$ are given by $\bar{\mu}(x,\theta)$ and $\bar{\sigma}^2(x,\theta)$, we may use the corresponding normal density

$$p_{QM}(\Delta, x, y, \theta) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2(x,\beta)}} \exp\left[-\frac{\big(y - x - \bar{\mu}(x,\alpha)\big)^2}{2\bar{\sigma}^2(x,\beta)} \right]$$

for the quasi-maximum likelihood estimation.

We may use

$$\mu_{M}(x,\alpha) = \Delta\mu(x,\alpha),$$
$$\sigma_{M}^2(x,\beta) = \Delta\sigma^2(x,\beta) + \frac{\Delta^2}{2}(\sigma\dot{\sigma})^2(x,\beta),$$

which is based on the Milstein approximation. Kessler (1997) uses

$$\mu_{K}(x,\theta) = \sum_{j=0}^{J} \frac{\Delta^j}{j!} L^j x$$
$$\sigma_{K}^2(x,\theta) = \Delta\sigma^2(x,\beta)$$
$$\left(1 + \frac{1}{\Delta\sigma^2(x,\beta)} \sum_{j=2}^{J} \Delta^j \sum_{k=0}^{J-j} \frac{\Delta^k}{k!} L^k \left(\sum_{a,b\geq 1, a+b=j} \frac{L^a x}{a!} \frac{L^b x}{b!} \right)\right),$$

where L is the infinitesimal generator.

Simulated ML Estimation

The exact transition may be obtained through simulations. As Gihman and Skorohod (1972) show, we have

$$p(\Delta, x, y, \theta) = \sqrt{\frac{\sigma(x,\beta)}{2\pi\Delta\sigma^3(y,\beta)}} \exp\left[-\frac{1}{2\Delta}\big(\tau(y,\beta) - \tau(x,\beta)\big)^2 + \int_x^y \frac{\mu(z,\alpha)}{\sigma^2(z,\beta)} dz \right]$$
$$\times \mathbb{E} \exp\left[\Delta \int_0^1 \omega\Big((1-t)\tau(x,\beta) + t\tau(y,\beta) + \sqrt{\Delta} W_t^\circ, \theta \Big) dt \right],$$

where $W_t^\circ = W_t - tW_1$ is Brownian bridge, τ is the Lamperti transformation and $\omega(x,\theta) = -(1/2)\big(\nu^2(x,\theta) + \dot{\nu}(x,\theta)\big)$ with

$$\nu(x,\theta) = \frac{\mu(\tau^{-1}(x,\beta),\alpha)}{\sigma(\tau^{-1}(x,\beta),\beta)} - \frac{1}{2}\dot{\sigma}(\tau^{-1}(x,\beta),\beta)$$

provided in particular that $|\omega(x,\theta)| = O(x^2)$ as $x \to \infty$. The expectation part involving Brownian bridge can be obtained by simulation up to arbitrary precision.

Utilizing the Chapman-Kolmogorov equation, Pedersen (1995) and Brandt and Santa-Clara (2002) suggest simulating the transition density with

$$p_N(\Delta, x, y, \theta) = \mathbb{E}\left[p^*\left(\frac{\Delta}{N}, X^*_{\Delta-\Delta/N}, y, \theta\right)\middle| X^*_0 = x\right],$$

where p^* is an approximated transition density based on, for example, the Euler approximation, and X^* is the corresponding process generated with that approximation. They show that p_N converges to the true transition density as $N \to \infty$, thus we can use it to obtain the exact ML estimation with arbitrary precision.

Continuous Time Asymptotics

Asymptotics for Continuous Time Process

To derive the limit distributions for the MLE's of diffusion models, we need to establish the asymptotics for continuous time processes

$$\int_0^T f(X_t)dt \qquad \text{and} \qquad \int_0^T g(X_t)dW_t$$

as $T \to \infty$, which we call, respectively, *the additive functional asymptotics* and *the martingale transform asymptotics*.

Function Classes

We consider classes of functions f such that

- f is integrable in m
- f is asymptotically homogeneous with limit function locally integrable in m

with speed density m, i.e., the classes of integrable functions and asymptotically homogeneous functions *in speed measure*.

Recurrence Property and Function Classes

For positive recurrent processes, $m(\mathcal{D})$ is finite and f is expected to be integrable in m for a large class of functions f. Therefore, we only consider the class of integrable functions in speed measure.

For null recurrent processes, $m(\mathcal{D}) = \infty$. Therefore, we consider the class of functions f such that f is integrable in m and the class of functions f such that f is only locally integrable in m, i.e., both classes of integrable and asymptotically homogeneous functions in speed measure.

Scale Transformation

We write

$$f(X_t) = (f \circ s^{-1})(s(X_t)) = f_s(X_t^s),$$

where s is the scale function. Recall that

$$ds(X_t) = (\sigma\dot{s})(X_t)dW_t = \left((\sigma\dot{s}) \circ s^{-1}\right)(s(X_t))dW_t.$$

Therefore, X^s is a driftless diffusion having speed measure

$$m_s(x) = \frac{2}{\sigma_s^2(x)}$$

with $\sigma_s = (\sigma\dot{s}) \circ s^{-1}$.

Integrability

We say that f is m-*integrable*, if f is integrable with respect to the speed measure m. Similarly, we say that f is m-*square integrable*, if $f \otimes f$ is integrable with respect to the speed measure m.

The scale transformation does not affect integrability. That is, f is m-integrable or m-square integrable if and only if f_s is m_s-integrable or m_s-square integrable.

Asymptotic Homogeneity

We say that f is m-*asymptotically homogeneous* if

$$f_s(\lambda x) \sim \kappa(f_s, \lambda) h(f_s, x)$$

for large λ, where $\kappa(f_s, \lambda)$ is nonsingular for all large λ and $h(f_s, \cdot)$ is locally integrable in both m_s and m_r. We call $\kappa(f_s, \cdot)$ and $h(f_s, \cdot)$ respectively the *asymptotic order* and *limit homogeneous function* of f. Similarly, we say that f is m-*square asymptotically homogeneous* if $h(f_s, \cdot)$ is locally square integrable in both m_s and m_r.

Regularity Condition for Null Recurrent Process

A null recurrent process is said to be *regular with index* $r > -1$ if for its speed density m_s in natural scale, we have

$$m_s(x) = m_r(x) + \varepsilon_r(x),$$

where m_r is a homogeneous function of degree $r > -1$, and ε_r is a locally integrable function such that $\varepsilon(x) = o(|x|^r)$ as $|x| \to \infty$.

Asymptotics for Positive Recurrent Process

If X is positive recurrent and f and $g \otimes g$ are m-integrable, then we have

$$\frac{1}{T}\int_0^T f(X_t)dt \to_{a.s.} \pi(f), \quad \frac{1}{\sqrt{T}}\int_0^T g(X_t)dW_t \to_d \mathbb{N}\big(0, \pi(gg')\big)$$

as $T \to \infty$.

For positive recurrent processes with integrability condition,

- The usual LLN and CLT hold respectively for the additive functional asymptotics and the martingale asymptotics
- The standard normal asymptotics apply

Limiting Null Recurrent Process

Let X be a regular null recurrent process with index $r > -1$ with speed density m_s in natural scale, and let the process $X^{sT})$ be defined on $[0, 1]$ for each T by $X_t^{sT} = T^{-1/(r+2)} X_{Tt}^s$. Then we have

$$(X^{sT}) \to_d (X^r)$$

as $T \to \infty$ in the space $\mathcal{C}[0, 1]$ of continuous functions defined on $[0, 1]$. Here X^r is defined by $X^r = W^r \circ \tau^r$ with

$$\tau_t^r = \inf \left\{ s \,\middle|\, \int_{\mathcal{D}} l_r(s, x) m_r(x) dx > t \right\}$$

for $0 \leq t \leq 1$, where W^r is standard Brownian motion and l_r is the local time of W^r.

Limit Process

The scaled limit process

$$\left[\left(\frac{a}{r+1}\right)^{\frac{1}{r+2}}+\left(\frac{b}{r+1}\right)^{\frac{1}{r+2}}\right]X^r$$

is a skew Bessel process in natural scale of dimension $2(r+1)/(r+2)$, where $m_r(x)=\Big(a1\{x\geq 0\}+b1\{x<0\}\Big)|x|^r$ with a and b are nonnegative constants such that $a+b>0$. It becomes

- Bessel process if $b=0$
- skew Brownian motion if $r=0$

Extension

If we define W^T, $W_t^T = T^{-1/2}W_{Tt}$, jointly with X^{sT}, then we have

$$\left(X^{sT}, W^T\right) \to_d \left(X^r, W\right)$$

jointly as $T \to \infty$, where W is Brownian motion identical to W^r for $r = 0$ and independent of W^r for all $r \neq 0$.

Remarks

If $r \neq 0$, the limit processes X^r and W are independent. This is because the Brownian motion driving X^{sT} becomes independent of W^T as $T \to \infty$ unless they run at the same speed

If $r = 0$, the Brownian motion defining the limit process X^r becomes identical to the limit process W. This case arises when the limit process X^r becomes a skew Brownian motion after scale transformation.

Asymptotics for Null Recurrent Process

Let X be null recurrent and regular with index $r > -1$.

(a) Under integrability condition, we have

$$\frac{1}{T^{1/(r+2)}} \int_0^T f(X_t)dt \to_d Km(f)A^{1/(r+2)}$$

$$\frac{1}{\sqrt{T^{1/(r+2)}}} \int_0^T g(X_t)dW_t \to_d \sqrt{K}m(gg')^{1/2}B \circ A^{1/(r+2)},$$

jointly as $T \to \infty$, where $A^{1/(r+2)}$ is the Mittag-Leffler process with index $1/(r+2)$ at time 1, and B is standard vector Brownian motion independent of $A^{1/(r+2)}$, and K is a constant depending upon r and the asymptotes of $m_s(x)/|x|^r$.

(b) Under asymptotic homogeneity condition, we have

$$\frac{1}{T}\kappa\big(f_s, T^{1/(r+2)}\big)^{-1}\int_0^T f(X_t)dt \to_d \int_0^1 h(f_s, X_t^r)dt$$

$$\frac{1}{\sqrt{T}}\kappa\big(g_s, T^{1/(r+2)}\big)^{-1}\int_0^T g(X_t)dW_t \to_d \int_0^1 h(g_s, X_t^r)dW_t$$

jointly as $T \to \infty$, where X^r is the limit process of X^s and W is identical to W^r for $r = 0$ and independent of W^r for all $r \neq 0$.

Remarks

Under required conditions, we have

$$T^{1/(r+2)} \ll T\kappa\left(f_s, T^{1/(r+2)}\right), \ \sqrt{T^{1/(r+2)}} \ll \sqrt{T}\kappa\left(f_s, T^{1/(r+2)}\right),$$

since, loosely put, asymptotically homogeneous functions are *bigger* than integrable functions.

The martingale transform asymptotics under integrability yield mixed normal limit distributions.

The martingale transform asymptotics under asymptotic homogeneity yield mixed normal limit distributions if $r \neq 0$. When $r = 0$, they yield Dickey-Fuller type limit distributions.

Asymptotic Theory of MLE

Basic Framework

If we let $\mathcal{S} = \partial\mathcal{L}/\partial\theta$ and $\mathcal{H} = \partial^2\mathcal{L}/\partial\theta\partial\theta'$, we have

$$\mathcal{S}(\hat{\theta}) = \mathcal{S}(\theta_0) + \mathcal{H}(\tilde{\theta})(\hat{\theta} - \theta_0),$$

where $\tilde{\theta}$ lies in the line segment connecting $\hat{\theta}$ and θ_0. To derive our asymptotics, it suffices to establish that

- AD1: $w^{-1}\mathcal{S}(\theta_0) \to_d N$,
- AD2: $w^{-1}\mathcal{H}(\theta_0)w^{-1\prime} \to_d M$ for some M positive definite a.s., and
- AD3: There is a sequence v such that $vw^{-1} \to 0$, and

$$\sup_{\theta\in\mathcal{N}} \left| v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1\prime} \right| \to_p 0,$$

  where $\mathcal{N} = \{\theta : |v'(\theta - \theta_0)| \leq 1\}$,

as $T \to \infty$ and $\Delta \to 0$ at appropriate rates.

Assumption 3.1

(a) $\sigma^2(x,\beta) > 0$,

(b) $\mu(x,\alpha)$, $\sigma^2(x,\beta)$ and $\ell(t,x,y,\theta)$ are infinitely differentiable in $t \geq 0$, $x, y \in \mathcal{D}$ and $\theta \in \Theta$, and that for any $f(t,x,y,\theta)$ of their derivatives we have $|f(t,x,y,\theta)| \leq g(x)$ for all $t \geq 0$ small, for all $y \in \mathcal{D}$ close to $x \in \mathcal{D}$ and for all $\theta \in \Theta$, where $g: \mathcal{D} \to \mathbb{R}$ is locally bounded and $|g(x)| \sim c|x|^p$ at boundaries $\pm\infty$ and $|g(x)| \sim c|x|^{-p}$ at boundary 0 for some constant $c > 0$,

(c) $\sup_{0 \leq t \leq T} |X_t| = O_p(T^q)$ if the boundaries are $\pm\infty$ and $(\inf_{0 \leq t \leq T} |X_t|)^{-1} = O_p(T^q)$ if one of the boundaries is 0, and

(d) $\Delta T^{4(pq+1)} \to 0$ as $T \to \infty$ and $\Delta \to 0$.

Lemma 3.1

Let $\ell(\Delta, x, y, \theta) = \Delta \log \left[\sqrt{\Delta} p(\Delta, x, y, \theta) \right]$ be the normalized log-likelihood for the transition density obtained by using any of the methods introduced earlier, and define the functional operators $\mathcal{A}$ and $\mathcal{B}$ as

$$\mathcal{A}f(t,x,y) = f_t(t,x,y) + \mu(y) f_y(t,x,y) + \frac{1}{2}\sigma^2(y) f_{yy}(t,x,y)$$

$$\mathcal{B}f(t,x,y) = \sigma(y) f_y(t,x,y).$$

Then we have

For all $x \in \mathcal{D}$ and θ in the interior of Θ,

$$\ell(0,x,x,\theta) = 0, \qquad \mathcal{A}\ell(0,x,x,\theta) = -\frac{\sigma^2(x)}{2\sigma^2(x,\beta)} - \log(\sigma(x,\beta)),$$

$$\mathcal{B}\ell(0,x,x,\theta) = 0, \qquad \mathcal{B}^2\ell(0,x,x,\theta) = -\frac{\sigma^2(x)}{\sigma^2(x,\beta)}$$

ignoring the terms which do not dependent upon θ, and

$$\mathcal{A}^2\ell(0,x,x,\theta) = -\frac{\mu^2(x,\alpha)}{\sigma^2(x,\beta)} + \mu(x)\frac{2\mu(x,\alpha)}{\sigma^2(x,\beta)} + \big(\sigma^2(x) - \sigma^2(x,\beta)\big)\ell_{tyy}(0,x,x,\theta),$$

$$\mathcal{AB}\ell(0,x,x,\theta) = \mathcal{BA}\ell(0,x,x,\theta) = \sigma(x)\frac{\mu(x,\alpha)}{\sigma^2(x,\beta)},$$

$$\mathcal{B}^3\ell(0,x,x,\theta) = 0$$

ignoring the terms which are independent of α.

Lemma 3.2

We have

$$\mathcal{S}_\alpha(\theta_0) = \int_0^T \frac{\mu'_\alpha}{\sigma}(X_t)dW_t + O_p(\sqrt{\Delta}T^{4pq+1})$$

$$\mathcal{S}_\beta(\theta_0) = \sqrt{\frac{2}{\Delta}}\int_0^T \frac{\sigma'_\beta}{\sigma}(X_t)dV_t + O_p(\Delta^{-1/4}T^{4pq+7/4})$$

and

$$\mathcal{H}_{\alpha\alpha}(\theta_0) = -\int_0^T \frac{\mu_\alpha\mu'_\alpha}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t)dW_t + O_p(\sqrt{\Delta}T^{4pq+1})$$

$$\mathcal{H}_{\alpha\beta}(\theta_0) = O_p(T^{3pq+1})$$

$$\mathcal{H}_{\beta\beta}(\theta_0) = -\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta\sigma'_\beta}{\sigma^2}(X_t)dt + O_p(\Delta^{-1/2}T^{3pq+1})$$

as $T \to \infty$ and $\Delta \to 0$.

Assumption 3.2

There exist positive sequences $w_\alpha(T)$ and $w_\beta(T)$ such that $w_\alpha(T), w_\beta(T) \to \infty$,

$$w_\alpha^{-1}(T) \int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt\, w_\alpha^{-1}(T) \to_d M_\alpha,$$

$$w_\beta^{-1}(T) \int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt\, w_\beta^{-1}(T) \to_d M_\beta$$

for some $M_\alpha, M_\beta > 0$ a.s. and

$$(w_\alpha \otimes w_\alpha)^{-1}(T) \int_0^T \frac{\mu_{\alpha\otimes\alpha}}{\sigma}(X_t)dW_t \to_p 0$$

as $T \to \infty$.

Assumption 3.3

If we let

$$f(x,\theta) = \mu(x)\left(\frac{\mu_{\alpha\otimes\alpha\otimes\alpha}}{\sigma^2}\right)(x,\theta) - \left(\frac{\mu\mu_{\alpha\otimes\alpha\otimes\alpha} + \mu_{\alpha\otimes\alpha}\otimes\mu_\alpha}{\sigma^2}\right)(x,\theta)$$

$$g(x,\theta) = \sigma(x)\left(\frac{\mu_{\alpha\otimes\alpha\otimes\alpha}}{\sigma^2}\right)(x,\theta),$$

there exists $\varepsilon > 0$ such that

$$T^{\varepsilon}(w_\alpha \otimes w_\alpha \otimes w_\alpha)^{-1} \sup_{\theta\in\mathcal{N}} \left| \int_0^T f(X_t,\theta)dt \right| \to_p 0$$

$$T^{\varepsilon}(w_\alpha \otimes w_\alpha \otimes w_\alpha)^{-1} \sup_{\theta\in\mathcal{N}} \left| \int_0^T g(X_t,\theta)dW_t \right| \to_p 0$$

as $T \to \infty$, where $\mathcal{N}$ is as defined as in AD3.

Remarks

Given our continuous time asymptotics, it is well expected that Assumptions 3.2 and 3.3 hold for a wide class of diffusion models including nonstationary, as well as stationary, processes.

Assumption 3.2 establish AD1 and AD2. Assumption 3.3 establish AD3. Therefore, the asymptotics of the MLE's follow readily from those of $\mathcal{S}(\theta_0)$ and $\mathcal{H}(\theta_0)$, which we obtain in Lemmas 3.1 and 3.2.

Primary Asymptotics

We have

$$\hat{\alpha} - \alpha \sim_p \left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t$$

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t,$$

where V is Brownian motion independent of W. Therefore, in particular, $\hat{\alpha}$ and $\hat{\beta}$ become uncorrelated as long as T is large and Δ is small, since V and (X, W) are independent.

Primary Asymptotics

We have

$$\hat{\alpha} - \alpha \sim_p \left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t$$

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t,$$

where V is Brownian motion independent of W. Therefore, in particular, $\hat{\alpha}$ and $\hat{\beta}$ become uncorrelated as long as T is large and Δ is small, since V and (X, W) are independent.

Primary Asymptotics for $\hat{\alpha}$

It follows from

$$\hat{\alpha} - \alpha \sim_p \left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t$$

that we have

- Consistency of $\hat{\alpha}$ requires $T \to \infty$
- Distribution of $\hat{\alpha}$ is generally non-normal unless $T \to \infty$

Primary Asymptotics for $\hat{\beta}$

It follows from

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t$$

that we have

- Consistency of $\hat{\beta}$ holds if either $T \to \infty$ or $\Delta \to 0$
- Distribution of $\hat{\beta}$ is mixed normal as long as $\Delta \to 0$ fast enough relative to $T \to \infty$

Limit Distributions: Positive Recurrent Case

For positive recurrent processes with integrability, we have

$$\sqrt{T}(\hat{\alpha} - \alpha) \to_d \mathbb{N}\left(0, \pi\left[\frac{\mu_\alpha \mu_\alpha'}{\sigma^2}\right]^{-1}\right)$$

$$\sqrt{\frac{T}{\Delta}}(\hat{\beta} - \beta) \to_d \mathbb{N}\left(0, \frac{1}{2}\pi\left[\frac{\sigma_\beta \sigma_\beta'}{\sigma^2}\right]^{-1}\right),$$

where $\pi(x) = m(x)/m(\mathcal{D})$ is the time-invariant stationary distribution of (X_t). The limit distributions are normal as in the standard stationary model.

Limit Distributions: Positive Recurrent Case

For positive recurrent processes with integrability, we have

$$\sqrt{T}(\hat{\alpha} - \alpha) \to_d \mathbb{N}\left(0, \mathbb{E}\left[\frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)\right]^{-1}\right)$$

$$\sqrt{\frac{T}{\Delta}}(\hat{\beta} - \beta) \to_d \mathbb{N}\left(0, \frac{1}{2}\mathbb{E}\left[\frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)\right]^{-1}\right),$$

where $\pi(x) = m(x)/m(\mathcal{D})$ is the time-invariant stationary distribution of (X_t). The limit distributions are normal as in the standard stationary model.

OU Process: $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta dW_t$

Primary Asymptotics

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta \left(\int_0^T \begin{pmatrix} 1 & X_t \\ X_t & X_t^2 \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} 1 \\ X_t \end{pmatrix} dW_t$$

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \beta \frac{V_T}{T}$$

Limit Distributions

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \to_d \mathbb{N} \left(0, \begin{bmatrix} \beta^2 - \frac{2\alpha_1^2}{\alpha_2} & -2\alpha_1 \\ -2\alpha_1 & -2\alpha_2 \end{bmatrix} \right)$$

$$\sqrt{T/\Delta}(\hat{\beta} - \beta) \to_d \mathbb{N}(0, \beta^2/2)$$

CEV Process: $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta_1 X_t^{\beta_2}$

Primary Asymptotics

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta_1 \left(\int_0^T \begin{pmatrix} X_t^{-2\beta_2} & X_t^{-2\beta_2+1} \\ X_t^{-2\beta_2+1} & X_t^{-2\beta_2+2} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} X_t^{-\beta_2} \\ X_t^{-\beta_2+1} \end{pmatrix} dW_t$$

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \sim_p \sqrt{\frac{\Delta}{2}} \beta_1 \left(\int_0^T \begin{pmatrix} 1 & \beta_1 \log X_t \\ \beta_1 \log X_t & \beta_1^2 \log^2 X_t \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} 1 \\ \beta_1 \log X_t \end{pmatrix} dV_t$$

Limit Distributions

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \to_d \mathbb{N} \left(0,\ \beta_1^2 \left[\mathbb{E} \begin{pmatrix} X_t^{-2\beta_2} & X_t^{-2\beta_2+1} \\ X_t^{-2\beta_2+1} & X_t^{-2\beta_2+2} \end{pmatrix} \right]^{-1} \right)$$

$$\sqrt{\frac{T}{\Delta}} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \to_d \mathbb{N} \left(0,\ \frac{\beta_1^2}{2} \left[\mathbb{E} \begin{pmatrix} 1 & \beta_1 \log(X_t) \\ \beta_1 \log(X_t) & \beta_1^2 \log^2(X_t) \end{pmatrix} \right]^{-1} \right)$$

Limit Distributions: NR/IN Case

For null recurrent processes with integrability, we have

$$\sqrt{T^{1/(r+2)}}(\hat{\alpha}-\alpha) \to_d \mathbb{MN}\left(0, \left[Km\left(\frac{\mu_\alpha \mu_\alpha'}{\sigma^2}\right)A^{1/(r+2)}\right]^{-1}\right)$$

$$\sqrt{\frac{T^{1/(r+2)}}{\Delta}}(\hat{\beta}-\beta) \to_d \mathbb{MN}\left(0, \frac{1}{2}\left[Km\left(\frac{\sigma_\beta \sigma_\beta'}{\sigma^2}\right)A^{1/(r+2)}\right]^{-1}\right)$$

independently, where K is a constant depending upon μ and σ, and $A^{1/(r+2)}$ is the Mittag-Leffler process with index $1/(r+2)$ at time 1. The convergence rate in T is

$$\sqrt{T^{1/(r+2)}} \ll \sqrt{T} \quad \text{with} \quad r > -1$$

and becomes slower than the positive recurrent case.

Limit Distributions: NR/IN Case

For null recurrent processes with integrability, we have

$$\sqrt{T^{1/(r+2)}}(\hat{\alpha}-\alpha) \to_d \mathbb{MN}\left(0, \left[Km\left(\frac{\mu_\alpha \mu_\alpha'}{\sigma^2}\right)A^{1/(r+2)}\right]^{-1}\right)$$

$$\sqrt{\frac{T^{1/(r+2)}}{\Delta}}(\hat{\beta}-\beta) \to_d \mathbb{MN}\left(0, \frac{1}{2}\left[Km\left(\frac{\sigma_\beta \sigma_\beta'}{\sigma^2}\right)A^{1/(r+2)}\right]^{-1}\right)$$

independently, where K is a constant depending upon μ and σ, and $A^{1/(r+2)}$ is the Mittag-Leffler process with index $1/(r+2)$ at time 1. The convergence rate in T is

$$\sqrt{T^{1/(r+2)}} \ll \sqrt{T} \quad \text{with} \quad r > -1$$

and becomes slower than the positive recurrent case.

Limit Distributions: NR/AH Case

For null recurrent processes with asymptotic homogeneity, we have

$$\sqrt{T}\kappa'\big(\nu_\alpha, T^{1/(r+2)}\big)(\hat{\alpha}-\alpha) \to_d \left(\int_0^1 hh'(\nu_\alpha, X_t^r)dt\right)^{-1}\int_0^1 h(\nu_\alpha, X_t^r)dW_t$$

$$\sqrt{\frac{T}{\Delta}}\kappa'\big(\tau_\beta, T^{1/(r+2)}\big)(\hat{\beta}-\beta) \to_d \frac{1}{\sqrt{2}}\left(\int_0^1 hh'(\tau_\beta, X_t^r)dt\right)^{-1}\int_0^1 h(\tau_\beta, X_t^r)dV_t$$

jointly with $\nu_\alpha = (\mu_\alpha/\sigma)\circ s^{-1}$ and $\tau_\alpha = (\sigma_\beta/\sigma)\circ s^{-1}$, where κ is asymptotic order and h is limit homogeneous function.

Remarks

The limit distribution of $\hat{\beta}$ is always mixed normal since X^r is independent of V

The limit distribution of $\hat{\alpha}$ is mixed normal if $r \neq 0$, in which case X^r is independent of W. In this case, $\hat{\alpha}$ and $\hat{\beta}$ become asymptotically independent.

The limit distribution of $\hat{\alpha}$ is essentially non-Gaussian and of unit-root type, if $r = 0$.

BOU: $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta dW_t$ with $\alpha_1 = \alpha_2 = 0$

Primary Asymptotics: Same as OU Process

Limit Distributions

$$\begin{pmatrix} \sqrt{T} & 0 \\ 0 & T \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \to_d \beta \left(\int_0^1 \begin{pmatrix} 1 & \beta W_t \\ \beta W_t & \beta^2 W_t^2 \end{pmatrix} dt \right)^{-1} \int_0^1 \begin{pmatrix} 1 \\ \beta W_t \end{pmatrix} dW_t$$

$$\sqrt{T/\Delta}(\hat{\beta} - \beta) \to_d \mathbb{N}(0, \beta^2/2)$$

In particular,

$$T(\hat{\alpha}_2 - \alpha_2) \to_d \left(\int_0^1 W_t^2 dt \right)^{-1} \int_0^1 W_t dW_t,$$

i.e., the Dickey-Fuller distribution which appears in the asymptotics of unit root tests.

NLD: $dX_t = (\alpha_1 + \alpha_2 X_t^{-1})dt + \sqrt{\beta_1 + \beta_2 X_t}dW_t$

Primary Asymptotics

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \left(\int_0^T \begin{pmatrix} \frac{1}{\beta_1+\beta_2 X_t} & \frac{1}{X_t(\beta_1+\beta_2 X_t)} \\ \frac{1}{X_t(\beta_1+\beta_2 X_t)} & \frac{1}{X_t^2(\beta_1+\beta_2 X_t)} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} \frac{1}{\sqrt{\beta_1+\beta_2 X_t}} \\ \frac{1}{X_t\sqrt{\beta_1+\beta_2 X_t}} \end{pmatrix} dW_t$$

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \begin{pmatrix} \frac{1}{4(\beta_1+\beta_2 X_t)^2} & \frac{X_t}{4(\beta_1+\beta_2 X_t)^2} \\ \frac{X_t}{4(\beta_1+\beta_2 X_t)^2} & \frac{X_t^2}{4(\beta_1+\beta_2 X_t)^2} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} \frac{1}{2(\beta_1+\beta_2 X_t)} \\ \frac{X_t}{2(\beta_1+\beta_2 X_t)} \end{pmatrix} dV_t$$

Limit Distributions

$$T^{\frac{1}{2}-\frac{\alpha_1}{\beta_2}} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \to_d \mathbb{MN}\left(0, \left[Km(f_\alpha f_\alpha')A^{1-2\alpha_1/\beta_2}\right]^{-1}\right)$$

$$\frac{T^{\frac{1}{2}-\frac{\alpha_1}{\beta_2}}}{\sqrt{\Delta}}(\hat{\beta}_1 - \beta_1) \to_d \mathbb{MN}\left(0, \left[2Km(f_{\beta_1}^2)A^{1-2\alpha_1/\beta_2}\right]^{-1}\right)$$

$$\sqrt{\frac{T}{\Delta}}(\hat{\beta}_2 - \beta_2) \to_d \mathbb{MN}\left(0, \left[2\int_0^1 \bar{f}_{\beta_2}^2(\bar{X}_t^s)dt\right]^{-1}\right)$$

for some functions f_α, f_{β_1} and $\bar{f}_{\beta_2}$ depending upon α and β

Simulation Results

Simulation Model

For simulations, we consider the CEV model

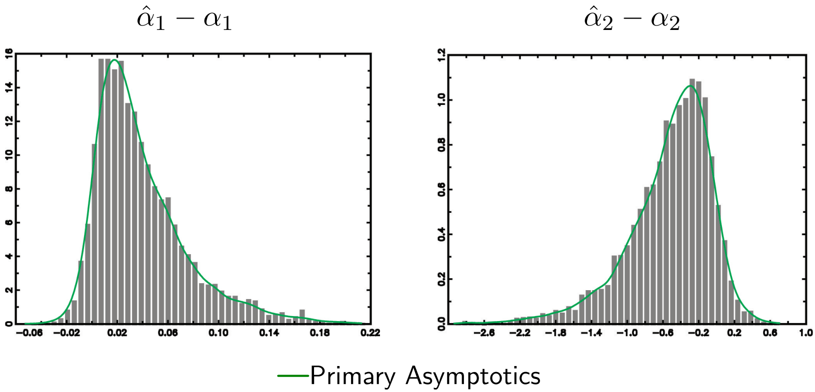
$$dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta_1 X_t^{\beta_2} dW_t$$

with parameter values $\alpha_1 = 0.0072$, $\alpha_2 = -0.09$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$, the estimated values for short rates in Aït-Sahalia (1999). The basic setup is

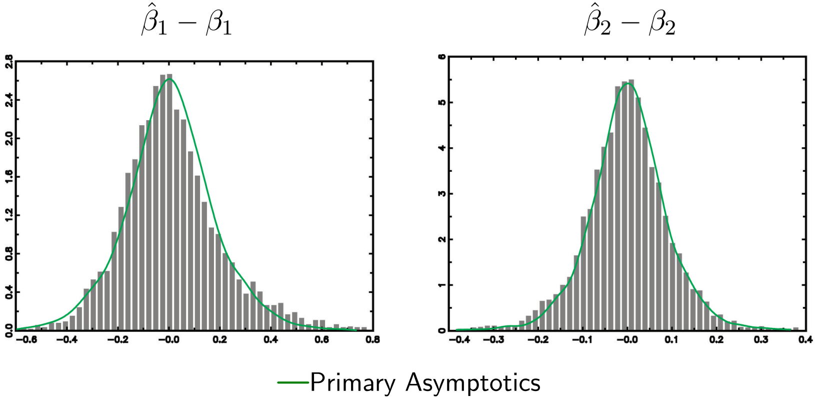
$$T = 10 \text{ years} \quad \text{and} \quad \Delta = 1 \text{ day}$$

unless specified otherwise. The simulations are based on 5,000 iterations.

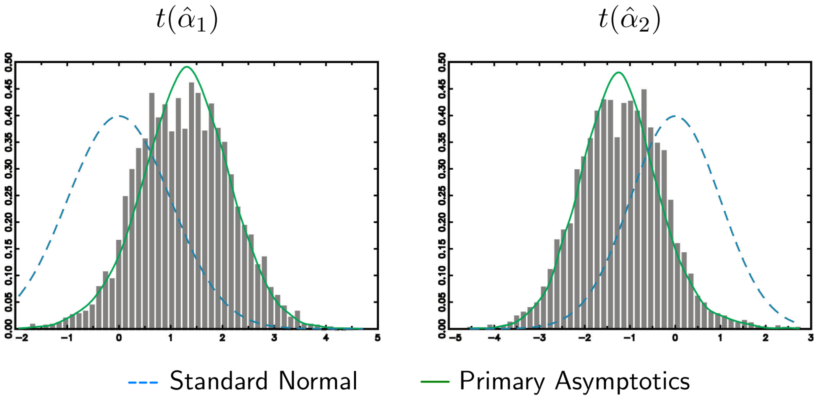
Drift Term Parameters



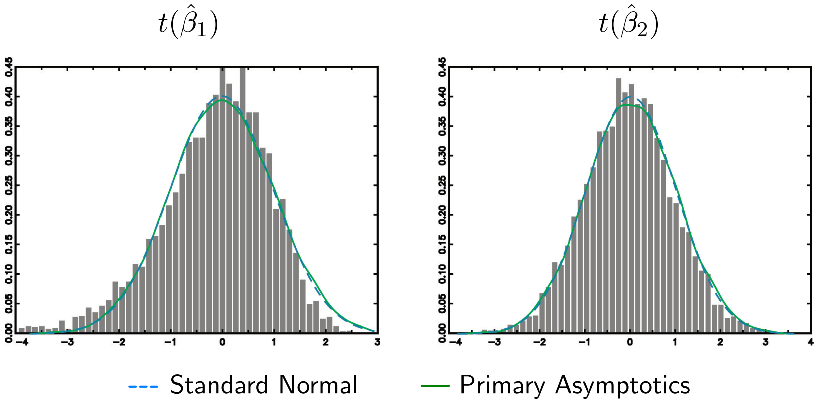
Diffusion Term Parameters



t-Statistics for Drift Term



t-Statistics for Diffusion Term



Bias Correction

10 Years/Daily

	Original		Corrected	
	α_1	α_2	α_1	α_2
Bias	0.04084	-0.55473	-0.00061	0.00264
(%)	(567.2)	(614.6)	(8.4)	(2.9)
SD	0.03551	0.44569	0.03590	0.45166
RMSE	0.05412	0.71159	0.03591	0.45167

50 Years/Daily

	Original		Corrected	
	α_1	α_2	α_1	α_2
Bias	0.00646	-0.10253	0.00004	-0.00080
(%)	(89.8)	(113.9)	(0.6)	(0.9)
SD	0.00776	0.12585	0.00762	0.12177
RMSE	0.01010	0.16233	0.00762	0.12177

Size Correction

10 Years/Daily

	Original		Corrected	
	α_1	α_2	α_1	α_2
1%	0.070	0.066	0.008	0.007
5%	0.202	0.189	0.052	0.052
10%	0.312	0.296	0.103	0.105

50 Years/Daily

	Original		Corrected	
	α_1	α_2	α_1	α_2
1%	0.029	0.028	0.012	0.013
5%	0.101	0.092	0.056	0.051
10%	0.177	0.166	0.105	0.099