The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1863-17

Advanced School and Conference on Statistics and Applied Probability in Life Sciences

24 September - 12 October, 2007

Analyzing trajectories: Functional predictors of univariate responses

Ian McKeague

Mailman School of Public Health
Columbia University
New York NY 10032, USA

Strada Costiera 11, 34014 Trieste, Italy - Tel. +39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it, www.ictp.it

Analyzing trajectories: functional predictors of univariate responses

Ian McKeague
Columbia University

October 1, 2007

Advanced School and Conference on Statistics and Applied Probability in Life Sciences, Trieste

Outline

National Collaborative Perinatal Project, 1959-1974

NIH study to investigate prenatal and familial antecedants of childhood growth and development, both physical and psychological.

Approximately 58,000 study pregnancies; mothers examined during pregnancy, labor, and delivery.

The children were given neonatal examinations and follow-up examinations at four, eight, and twelve months, and three, four, seven, and eight years.

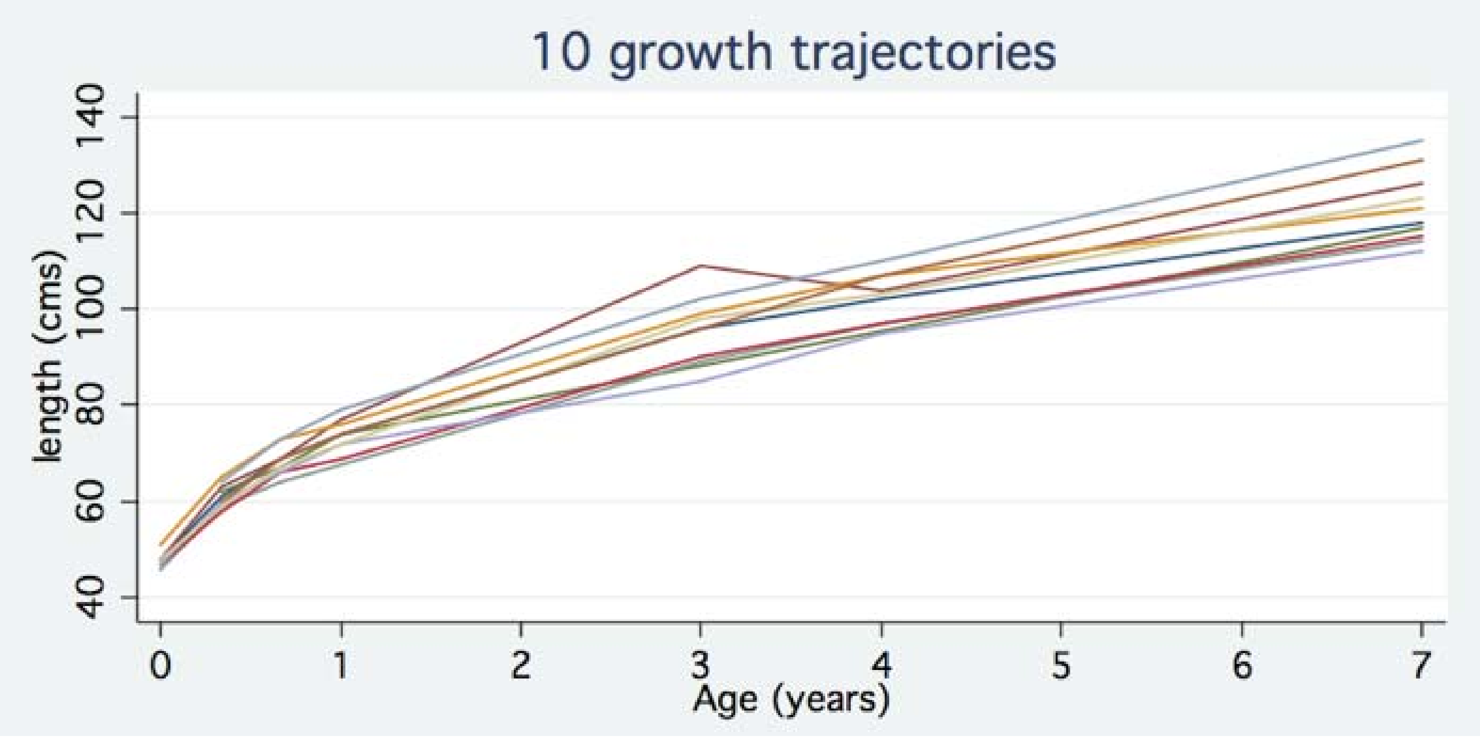
Motivating question

Role of early life in chronic disease and cognitive development:

- birth size negatively associated with cardiovascular morbidity and mortality
- birth size positively related to cancer
- birth size positively associated with cognitive ability

Is there a **sensitive period** during which growth rate is predictive of cognitive ability?

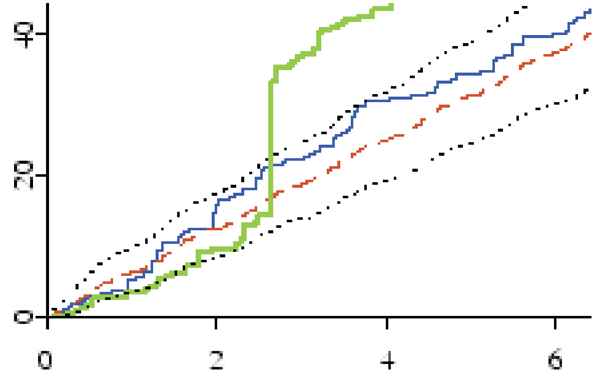
Examples of growth trajectories



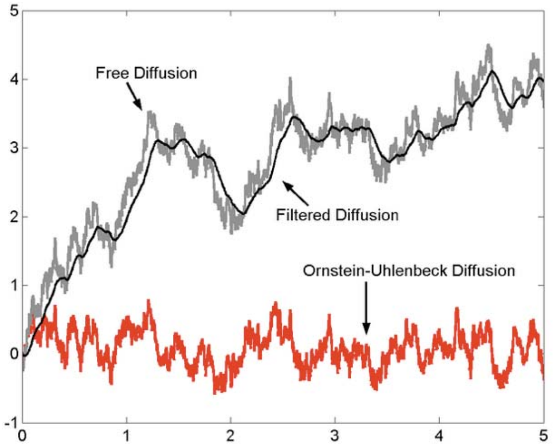
Linearly interpolated NCPP height data

Models for trajectories

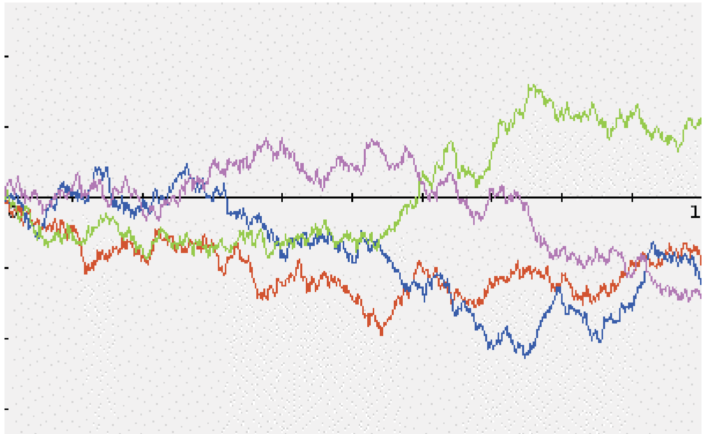
Compound Poisson processes



Diffusion processes



Brownian motion



Why should we model the trajectories?

- We are interested in using the trajectories as predictors without first reducing them finite-dimensional vectors.
- For smooth trajectories (Lipschitz in time), “standard” regression methods are available: $\sqrt{n}$-rates, asymptotic normality, robust estimates of standard errors, bootstrap works, etc.
- For “rough” trajectories (with jump discontinuities or unbounded variation), standard methods will not work.

Inference for sensitive periods

An interpretable ("working") model:

$$Y = \alpha + \beta X(\theta) + Z^T\gamma + \epsilon$$

Scalar response: Y = IQ at age 7

Predictor of interest: $X(\theta)$ = growth rate at time θ

Other covariates: Z = (birth weight, gestational age, ...)

ϵ has mean zero, finite variance σ^2, independent of (X, Z).

Least squares estimates:

$$(\hat{\theta}_n, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n) = \mathrm{argmin}_{(\theta,\alpha,\beta,\gamma)} \sum_{i=1}^{n} [Y_i - \alpha - \beta X_i(\theta) - Z_i^T\gamma]^2$$

Functional linear regression

$$Y = \alpha + \int_0^1 f(t)X(t)\,dt + Z^T\gamma + \epsilon$$

- Ramsay and Silverman (1997, 2002) popularized the method of functional principal components for nonparametric estimation of f.
- Hall and Horowitz (2007) showed such estimators achieve the minimax rate (in terms of the IMSE). If X is Brownian motion, minimax rate is between $n^{1/4}$ and $n^{1/2}$.

Spectroscopy application: concentration Y of a protein regressed on intensity $X(t)$ of reflected radiation at wavelength t.

Interpretable functional regression

Working model:

$$Y = \alpha + \beta X(\theta) + \epsilon.$$

Least squares estimator:

$$(\hat{\theta}_n, \hat{\alpha}_n, \hat{\beta}_n) = \operatorname{argmin}_{(\theta,\alpha,\beta)} \sum_{i=1}^{n} [Y_i - \alpha - \beta X_i(\theta)]^2$$

estimates

$$(\theta_0, \alpha_0, \beta_0) = \operatorname{argmin}_{(\theta,\alpha,\beta)} E[Y - \alpha - \beta X(\theta)]^2.$$

Key Question: Is there a rate of convergence r_n such that $r_n(\hat{\theta}_n - \theta_0)$ has a non-degenerate limiting distribution?

Two cases: 1) true working model, 2) misspecified working model, with the data satisfying a functional linear model:

$$Y = \int_0^1 f(t)X(t)\,dt + \epsilon.$$

Change-point estimation

Single-jump process: $X(t) = 1\{T \geq t\}$

Correctly specified case

$$Y = \alpha + \beta X(\theta) + \epsilon$$

$$n(\hat{\theta}_n - \theta_0) \to_d \text{minimizer of a compound Poisson process}$$
$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to_d \text{normal}$$

Koul, Qian and Surgailis (2003): two-phase linear regression.

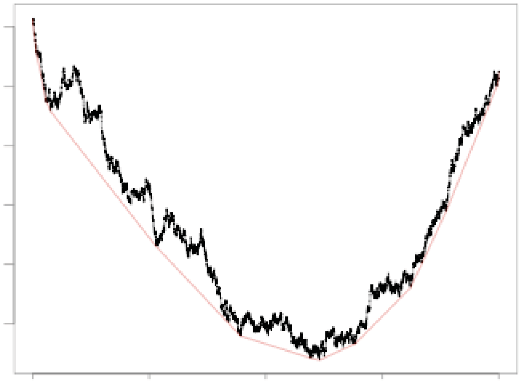
Misspecified case

$$Y = F(T) + \epsilon \qquad \text{where } F' = f$$

Banerjee and McKeague (2007): split point estimation.

$$n^{1/3}(\hat{\theta}_n - \theta_0) \to_d \text{ scaled Chernoff}$$
$$n^{1/3}(\hat{\beta}_n - \beta_0) \to_d \text{ scaled Chernoff}$$

Minimizer of Brownian motion with drift



Parabolic drift: Groeneboom (1985) showed that

$$\operatorname{argmin}_{t\in\mathbb{R}}(B(t)+t^2)$$

has a density that can be expressed in terms of zeros of the Airy function. Known as the Chernoff distribution.

Triangular drift: Bhattacharya and Brockwell (1976) showed that

$$\operatorname{argmin}_{t\in\mathbb{R}}(B(t)+|t|)$$

has a density that can be expressed in terms of Φ.

Asymptotic theory for M-estimators

Consider the general M-estimator

$$\hat{\theta}_n = \mathrm{argmin}_\theta \mathbb{M}_n(\theta)$$

of $\theta_0 = \mathrm{argmin}_\theta \mathbb{M}(\theta)$, where $\mathbb{M}(\theta) = E[m_\theta]$ and

$$\mathbb{M}_n(\theta) = \mathbb{P}_n[m_\theta] = \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i, Y_i).$$

Assumption: there is a metric d on Θ such that

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \gtrsim d^2(\theta, \theta_0)$$

for all θ in a neighborhood of θ_0.

Brownian trajectories

Working model: $Y = X(\theta) + \epsilon$

$X(t)$ is Brownian motion, $m_\theta(X, Y) = (Y - X(\theta))^2$

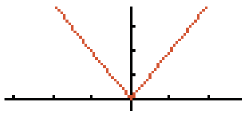
Misspecified case: data from the functional linear model

$$Y = \int_0^1 f(t)X(t)\,dt + \epsilon.$$

Easy to show that $\mathbb{M}(\theta) = E[m_\theta]$ is twice differentiable, so d is Euclidean distance: $d(\theta, \theta_0) = |\theta - \theta_0|$.

Correctly specified case: data from $Y = X(\theta_0) + \epsilon$. Now

$$\mathbb{M}(\theta) = E[X(\theta) - X(\theta_0)]^2 + \sigma^2 = |\theta - \theta_0| + \sigma^2.$$



$\mathbb{M}$ is *not differentiable* at θ_0, and $d(\theta, \theta_0) = \sqrt{|\theta - \theta_0|}$.

Rate of convergence

A lower bound on the rate of convergence r_n can be found in terms of the continuity modulus

$$w_n(\delta) = \sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})|,$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ is the empirical process.

Theorem (van der Vaart and Wellner). If $E[w_n(\delta)] \lesssim \delta^\alpha$ for some $0 < \alpha < 2$, then

$$n^{1/(4-2\alpha)} d(\hat{\theta}_n, \theta_0) = O_p(1).$$

Example: $\alpha = 1$ gives the "usual" rate $n^{1/2}$.

Example: $\alpha = \frac{1}{2}$ gives rate $n^{1/3}$.

Key steps

A result from empirical process theory (Pollard, 1989) gives

$$E[w_n(\delta)] \le J_{[\,]}(1, \mathcal{M}_\delta)\{EM_\delta^2\}^{1/2}.$$

$J_{[\,]}(1, \mathcal{M}_\delta)$ is the bracketing entropy integral of the class of functions

$$\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}.$$

M_δ is an envelope function for $\mathcal{M}_\delta$.

Brownian trajectories are Lipschitz: for $0 < \alpha < 1/2$,

$$|X(t) - X(s)| \le K|t - s|^\alpha \quad \forall\ t, s \in [0, 1]$$

where K has moments of all orders [Kolmogorov's continuity theorem].

Lemma: m_θ is "Lipschitz in parameter":

$$|m_{\theta_1} - m_{\theta_2}| \le L|\theta_1 - \theta_2|^\alpha, \quad \text{where } EL^2 < \infty.$$

Corollary: $J_{[\,]}(1, \mathcal{M}_\delta) < \infty$.

Envelope function

Self-similarity of the Brownian trajectories is used to bound the second moment of the continuity modulus

$$F_\delta = \sup_{|\theta-\theta_0|<\delta} |m_\theta - m_{\theta_0}|.$$

Self-similarity: $X(\delta t) =_d \delta^{1/2} X(t)$.

$$\{EF_\delta^2\}^{1/2} \lesssim \left\{E \sup_{|\theta-\theta_0|<\delta} |X(\theta) - X(\theta_0)|^4\right\}^{1/4} \lesssim \sqrt{\delta}.$$

Results

Correctly specified case

$d(\theta, \theta_0) = \sqrt{|\theta - \theta_0|}$, envelope function $M_\delta = F_{\delta^2}$. Get the "usual" rate $n^{1/2}$ with respect to d, which translates to rate n with respect to Euclidean metric:

$$n(\hat{\theta}_n - \theta_0) \to_d \mathrm{argmin}_{t \in \mathbb{R}}(2\sigma B(t) + |t|),$$

where B is a two-sided Brownian motion.

Misspecified case

$d(\theta, \theta_0) = |\theta - \theta_0|$, envelope function $M_\delta = F_\delta$. Cube-root rate:

$$n^{1/3}(\hat{\theta}_n - \theta_0) \to_d \mathrm{argmin}_{t \in \mathbb{R}}(2aB(t) + bt^2)$$

and a scaled Chernoff limit, as in change-point estimation.

Full model: $Y = \alpha + \beta X(\theta) + \epsilon$, LS estimators of α_0, β_0 have $\sqrt{n}$ and $n^{1/3}$ rates for the correctly specified and misspecified cases, respectively.

Details

Idea is to localize the criterion function:

$$\widetilde{\mathbb{M}}_n(h) = s_n[\mathbb{M}_n(\theta_0 + h/r_n) - \mathbb{M}_n(\theta_0)]$$

$$r_n(\hat{\theta}_n - \theta_0) = \hat{h}_n = \mathrm{argmin}_{h\in\mathbb{R}}\widetilde{\mathbb{M}}_n(h)$$

Need to adjust the scaling s_n so we can apply the

Argmin continuous mapping theorem: If $\widetilde{\mathbb{M}}_n \to_d \widetilde{\mathbb{M}}$ in $B_{\mathrm{loc}}(\mathbb{R})$ and $\hat{h}_n = O_p(1)$, then

$$\hat{h}_n \to_d \mathrm{argmin}_h\widetilde{\mathbb{M}}(h)$$

Details (cont'd)

Correctly specified case: $s_n = r_n = n$

$$\begin{aligned}\widetilde{\mathbb{M}}_n(h) &= n(\mathbb{P}_n - P)(m_{\theta_0+h/n} - m_{\theta_0}) + nP(m_{\theta_0+h/n} - m_{\theta_0}) \\ &= n^{-1/2}\mathbb{G}_n[Z_n(h)^2] - 2\mathbb{G}_n[\epsilon Z_n(h)] + |h|,\end{aligned}$$

where $Z_n(h) \equiv \sqrt{n}[X(\theta_0 + h/n) - X(\theta_0)]$, and first term is $o_p(1)$.

$Z_n(h) =_d B(h)$ as processes on the real line, so

$$\mathbb{G}_n[\epsilon Z_n(h)] =_d B(h)\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\right)^{1/2} \to_d \sigma B(h)$$

Conclude $\widetilde{\mathbb{M}}_n(h) \to_d 2\sigma B(h) + |h|$ in $B_{\mathrm{loc}}(\mathbb{R})$.

Fractional Brownian motion

Gaussian process $X(t)$, $t \in \mathbb{R}$, mean zero, covariance

$$\mathrm{Cov}\{X(t), X(s)\} = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}),$$

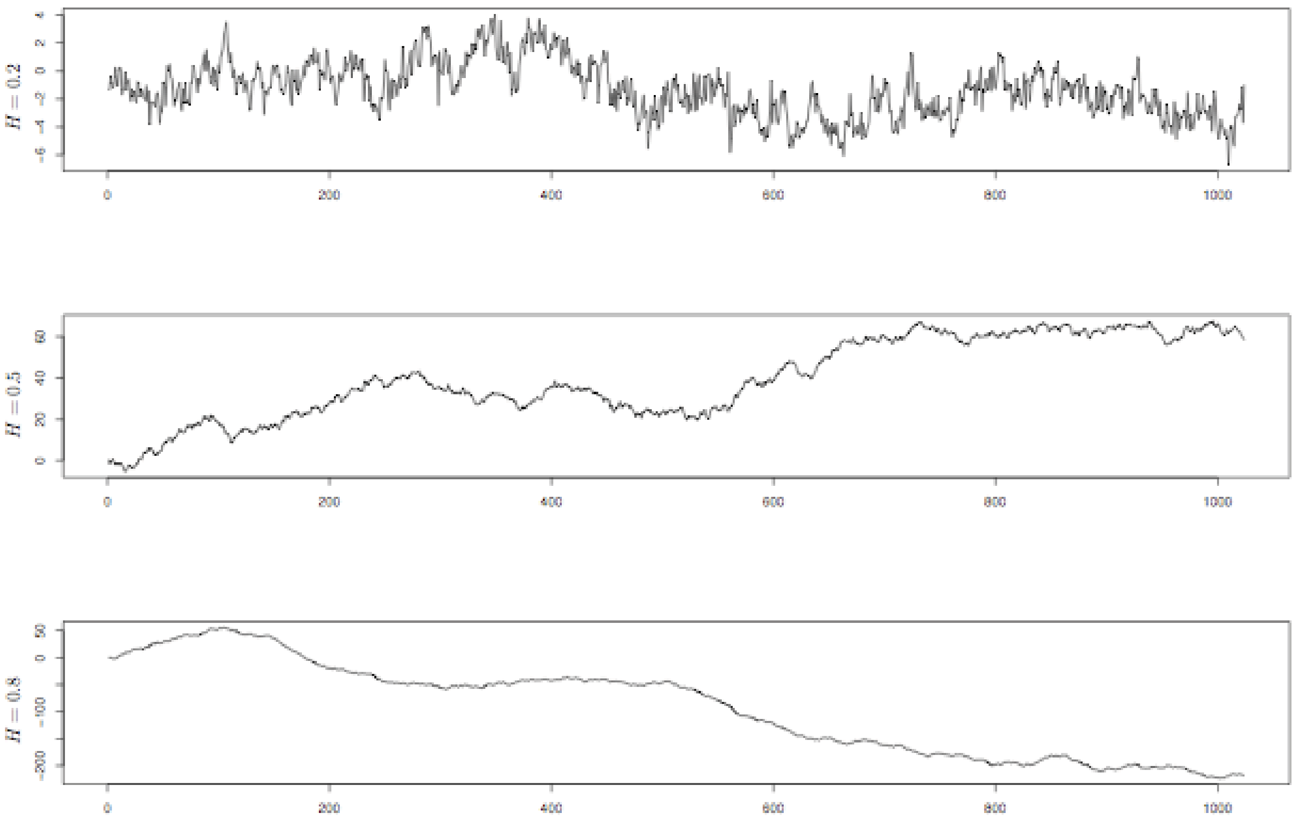
$H \in (0,1]$ is the **Hurst exponent**.

- $H = 1/2$ gives two-sided Brownian motion
- $H = 1$ gives a straight line: $X(t) = tZ$ where $Z \sim N(0,1)$.
- self-similarity: $X(\delta t) =_d \delta^H X(t)$ for all $\delta > 0$
- trajectories are locally Lipschitz of order $\alpha < H$:

$$|X(t) - X(s)| \leq K|t-s|^{\alpha} \quad \forall\ t, s \in [0,1]$$

where K has moments of all orders.

fBm trajectories



R function `fbmSim` used for simulation of fBm

Results

Correctly specified case:

$$n^{1/(2H)}(\hat{\theta}_n - \theta_0) \to_d \mathrm{argmin}_{t\in\mathbb{R}}(2\sigma B_H(t) + |t|^{2H}).$$

Rate becomes *arbitrarily fast* as $H \to 0$.

Misspecified case:

$$n^{1/(4-2H)}(\hat{\theta}_n - \theta_0) \to_d \mathrm{argmin}_{t\in\mathbb{R}}(2aB_H(t) + bt^2)$$

Rate becomes *slower* as H decreases — as slow as $n^{1/4}$.

Partial misspecification

$$Y = \alpha + \beta X(\theta) + \int_0^1 f(t)X(t)\,dt + \epsilon.$$

If $H \leq 1/2$ and $\int |f|$ is sufficiently small, then θ_0 coincides with the true θ, and

$$n^{1/(2H)}(\hat{\theta}_n - \theta_0) \to_d \operatorname{argmin}_{t\in\mathbb{R}}(2aB_H(t) + |t|^{2H}),$$

where

$$a^2 = \sigma^2 + E\left(\int_0^1 f(t)X(t)\,dt\right)^2.$$

CIs in the correctly specified case

$100(1-\alpha)\%$ confidence interval for θ_0:

$$\hat{\theta}_n \pm \left(\frac{\sigma}{\sqrt{n}}\right)^{1/H} z_{H,\alpha/2}$$

where $z_{H,\alpha}$ is the upper α-quantile of

$$Z_H = \operatorname{argmin}_{t\in\mathbb{R}}(B_H(t) + |t|^{2H}/2).$$

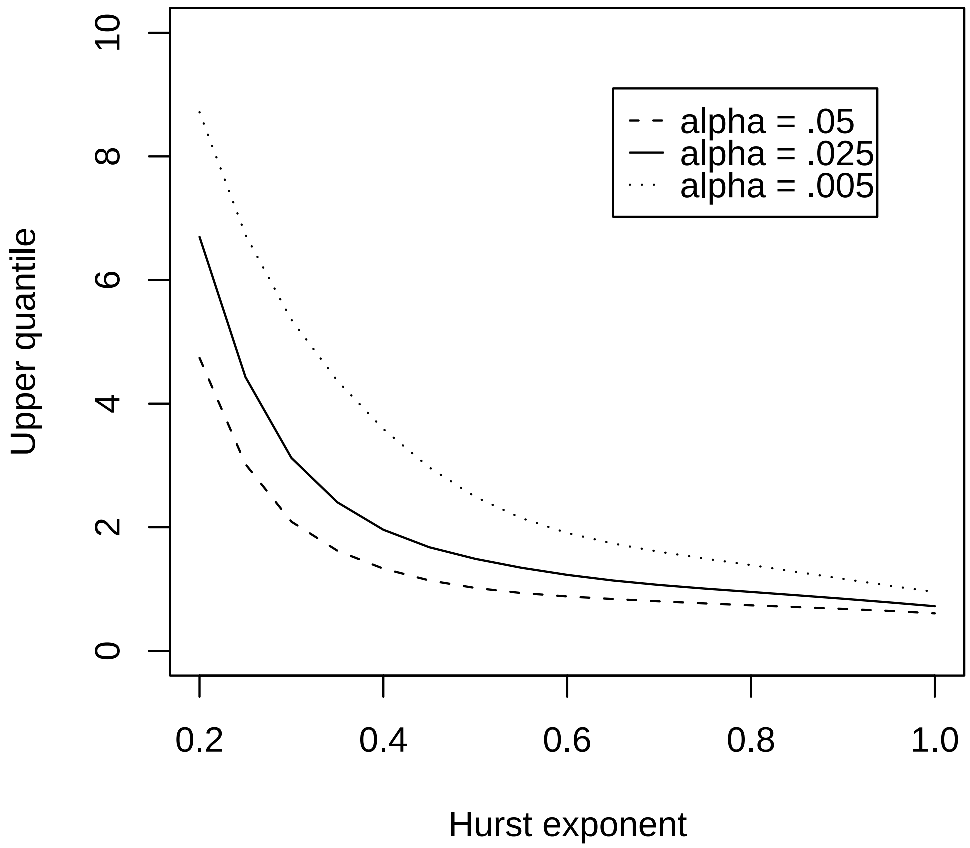
Full model: $Y = \alpha + \beta X(\theta) + \epsilon$

$$\hat{\theta}_n \pm \left(\frac{\sigma}{\hat{\beta}_n \hat{\gamma}_n \sqrt{n}}\right)^{1/H} z_{H,\alpha/2}$$

given $X(t) = X_0 + \gamma \tilde{X}(t)$ with $\tilde{X}(t)$ a standard fBm.

Quantiles of Z_H

$Z_H^* = \exp(-1/H) Z_H$ has upper quantiles given by:



Simulation examples

Correctly specified case:

$$Y = \alpha + \beta X(\theta) + \epsilon,$$

where $\alpha = 0$, $\beta = 1$, $\theta_0 = 1/2$, $\epsilon \sim N(0, .25)$, $n = 20$.

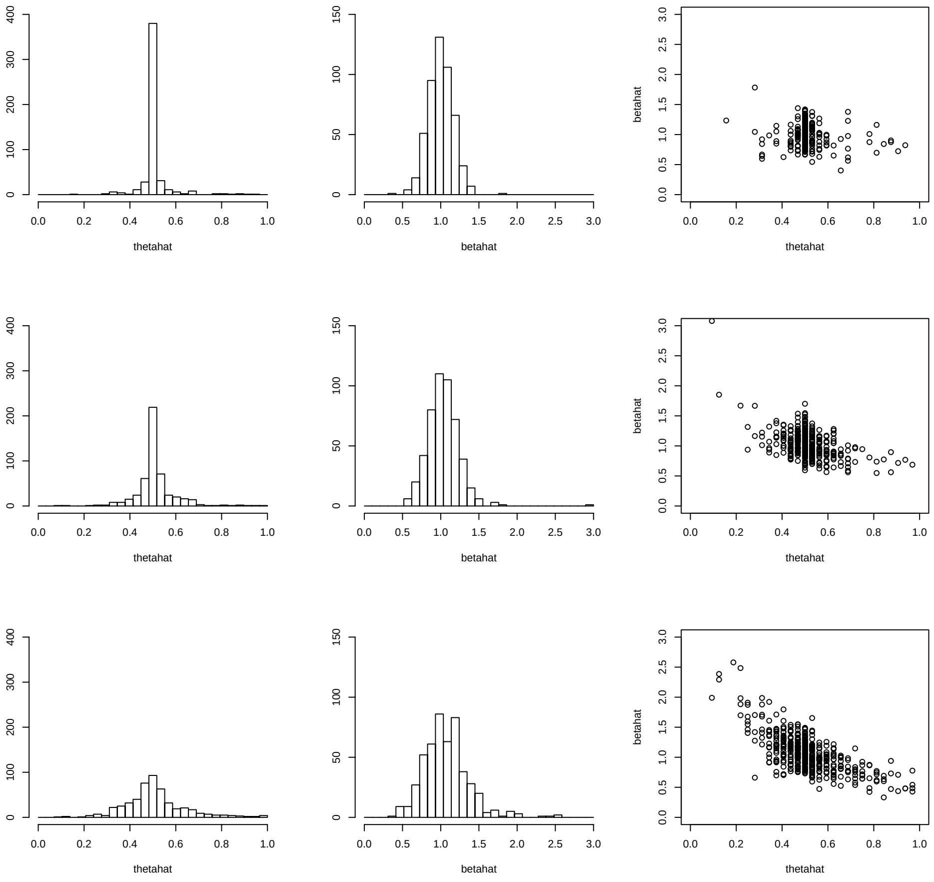
Partially misspecified case:

$$Y = \alpha + \beta X(\theta) + \int_0^1 f(t)X(t)\,dt + \epsilon,$$

where $f(t) = 1/2$ and true $\theta = 1/2$.

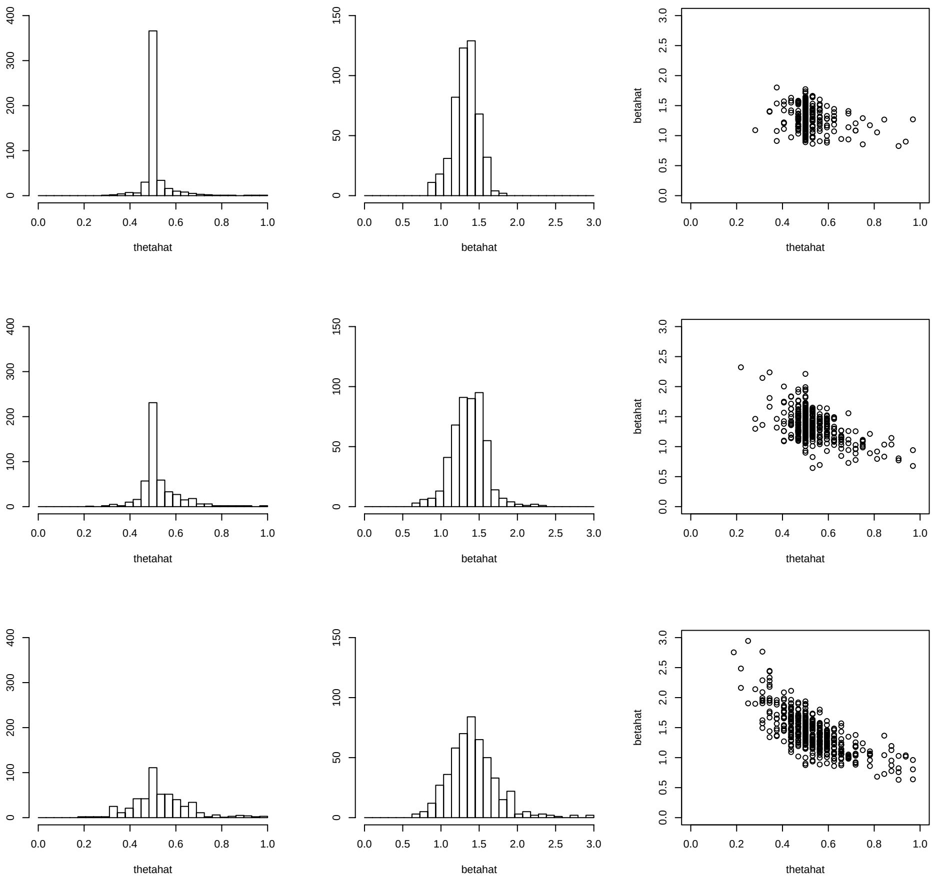
Hurst exponent: $H = .3, .5$ and $.7$

Correctly specified case



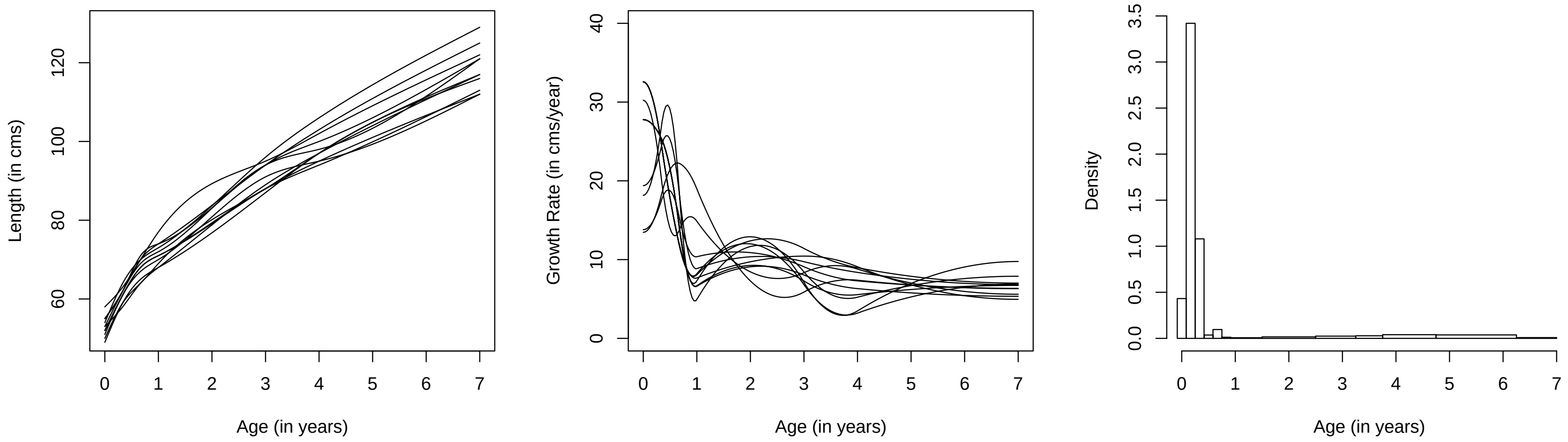
$H = .3$ (top), $H = .5$ (middle), and $H = .7$ (bottom), based on 500 samples of size $n = 20$. CI widths: 0.12, 0.27 and 0.38, respectively.

Partially misspecified case



$H = .3$ (top), $H = .5$ (middle), and $H = .7$ (bottom)

Application to growth curves



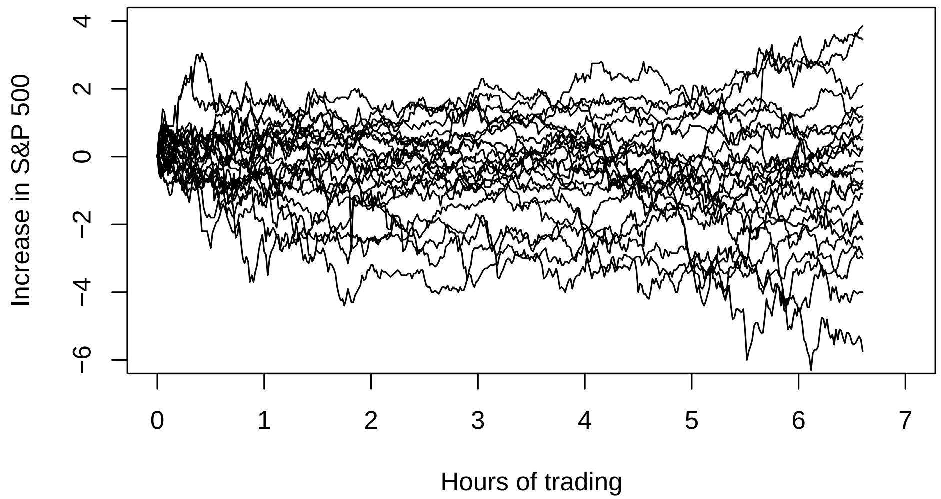
NCPP growth curves based on natural cubic spline interpolation between the observation times (left), corresponding growth rate trajectories (middle), and histogram of $\hat{\theta}_m$ for 500 subsamples of size $m = 500$ (right).

$n = 5704$, $\hat{\theta}_n = 2$ months.

Application to NYSE data

Black–Scholes model of stock prices: $H = 1/2$

$X(t)$ = increase in S. & P. 500-stock index over trading day;
Y = total increase over next day



$n = 23$ trading days (from August 1995)

95% CI for θ_0: 0–57 minutes after the opening bell

Conclusion

- Introduced "interpretable" functional linear regression models with fBm trajectories as predictors.
- Derived confidence intervals for sensitive time points in terms of the Hurst exponent.
- Feasible extensions:
  - multiple time points (model selection issues arise)
  - diffusion processes (rates as for Brownian motion)
  - Lévy processes (stationary independent increments)
  - multiparameter fBm
  - Cox regression: $\lambda(t|X) = \lambda_0(t)\exp(\beta X(\theta))$, $\mathcal{F}_0$-measurable X.