Long Memory

Letting $\gamma_j = E(u_t u_{t-j})$, weak dependence (short memory) of a stationary process $\{u_t\}$ is where

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty.$$

This condition is satisfied by any ARMA(p, q) process with stable roots. In that case can show that $|\gamma_j| = O(\mu_{\max}^j)$. This is called *exponential memory decay*.

The borderline case is $|\gamma_j| = O(j^{-1})$, since $\sum_{j=1}^{m} j^{-1} = O(\log m)$, whereas $\sum_{j=1}^{\infty} j^{-1-\varepsilon} < \infty$ for any $\varepsilon > 0$. With long memory, the *square-root rule* for the variation of partial sums is violated:

- If $\{u_t\}$ is weakly dependent, with mean 0 and variance σ^2 then

$$\sum_{t=1}^{T} u_t = O_p(\sqrt{T})$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} N(0, \sigma^2)$$

- If $\{u_t\}$ is long memory with mean 0 and variance σ^2 then

$$\sum_{t=1}^{T} u_t = O_p(T^H)$$

for *Hurst exponent* $H > \frac{1}{2}$, $|\gamma_j| = O(j^{2(H-1)})$ and

$$\sum_{j=0}^{m} |\gamma_j| = O(m^{2H-1}).$$

The usual Central Limit Theorem does not apply here.

Spectral Analysis

The spectrum (spectral density) of a stationary process is a function on $[-\pi, \pi]$ defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\lambda}$$
$$= \frac{\gamma_0}{2\pi} + \frac{1}{\pi} \sum_{j=1}^{\infty} \gamma_j \cos(\lambda j) \geq 0 \quad -\pi \leq \lambda \leq \pi$$

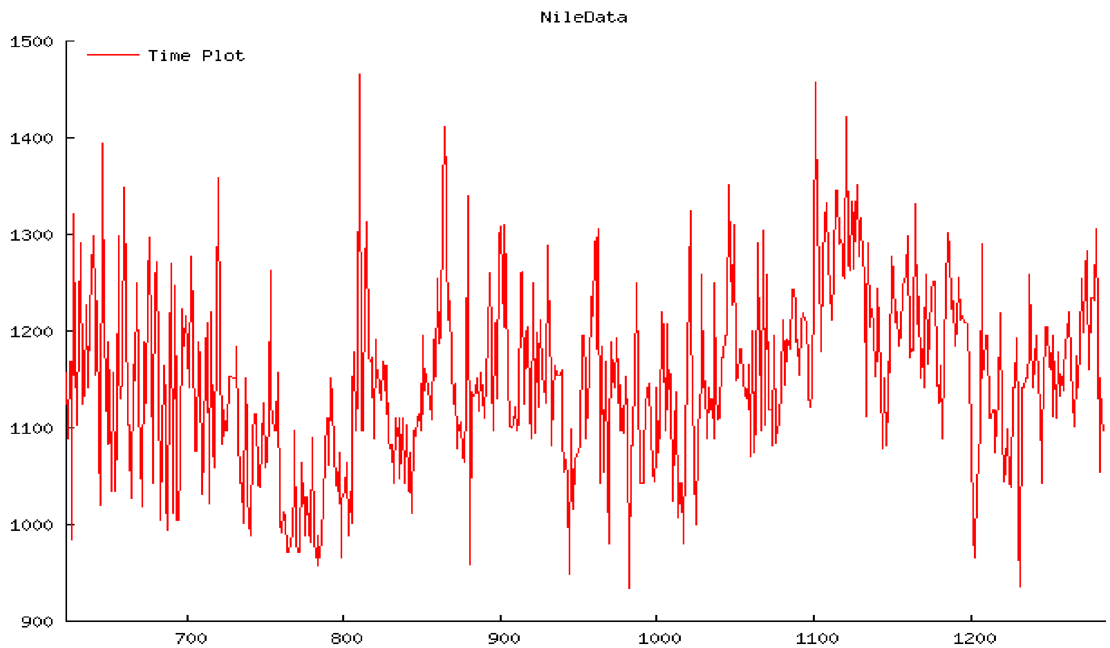
Represents the amount of variation at each frequency.

- An iid process is "white noise",
$$f(\lambda) = \frac{\gamma_0}{2\pi}$$
  at all λ (flat spectrum).
- A weakly dependent process has $0 < f(0) < \infty$.
- A long memory process has $f(0) = \infty$, and $f(\lambda) = O(|\lambda|^{1-2H})$ as $\lambda \to 0$.
- The case $f(0) = 0$ is called "anti-persistence".
  - The best known case is an "over-differenced" process,
$$v_t = u_t - u_{t-1}$$
    where $u_t \sim$ iid, Note that $\gamma_0 = 2\sigma^2$, $\gamma_1 = \gamma_{-1} = -\sigma^2$, and $\gamma_{|j|} = 0$ for $j > 1$.

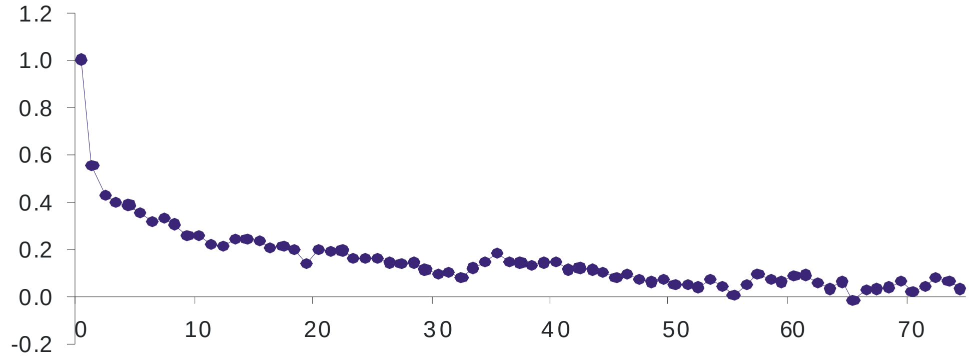
Example: Nile River Data

Harold Hurst was a hydrologist employed on the Aswan dam project on the Nile

He analysed a well-known data set, 600 years of unbroken observations (622-1284AD) on the annual minimum height of the Nile near Cairo:



Here are the first 70 autocorrelations:



These decrease too slowly to represent a weakly dependent process.

Testing for Long Memory

Hurst proposed a test of the hypothesis $H = \frac{1}{2}$, the *rescaled range* (RS) test.

Letting $S_t = \sum_{s-1}^{t}(u_s - \bar{u})$,

$$\mathrm{RS} = \frac{1}{\sqrt{T}\,\hat{\sigma}}\left(\max_{1\leq t\leq T}\{S_t\} - \min_{1\leq t\leq T}\{S_t\}\right)$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(u_t - \bar{u})^2$.

- If $H = \frac{1}{2}$, this statistic has a known large-sample distribution that can be tabulated by simulation.
- Can test $H > \frac{1}{2}$ from tabulation of upper tail, as well as $H < \frac{1}{2}$ with tabulation of lower tail.
- Strictly, this test only valid for cases where $\{u_t\}$ i.i.d. on null hypothesis, not just weakly dependent.

Andrew Lo (1991) proposed a modified test which replaces $\hat{\sigma}^2$ by a HAC consistent estimator,

$$\hat{\omega}^2 = \frac{1}{T}\left(\sum_{t=1}^{T} \varepsilon_t^2 + 2\sum_{j=1}^{M_T} w_T(j) \sum_{t=1}^{T-j} \varepsilon_t \varepsilon_{t+j}\right)$$

(Newey and West 1987) where

- $w_T(j) = 1 - j/M_T$ (Bartlett kernel)
- $M_T = O(T^{1/3})$.

Can be shown that RS $\sim T^{H-1/2}$, and accordingly that

$$\tfrac{1}{2} + \log(\mathrm{RS})/\log T \to H$$

Lo's RS statistic for Nile data is 3.013 (p-value < 0.005).

Note, $\frac{1}{2} + \log(3.013)/\log(664) = 0.67$

Estimation of the long memory parameter

Define $d = H - \frac{1}{2}$.

- $d = 0$ represents short memory,
- $d > 0$ is long memory,
- $d < 0$ is anti-persistence.

Assume that the spectral density takes the form

$$f(\lambda) = |\lambda|^{-2d} g(\lambda)$$

where $0 < g(0) < \infty$, $g'(0) = 0$ and $g''(0) < \infty$.

This pattern is shared by a variety of known stationary time series. (see later).

The spectrum of a process $\{x_t\}$ can be estimated from a sample $x_1, \ldots, x_T$ by the *periodogram*

- This is the squared modulus of the discrete Fourier transform of the series.

$$I(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{i\lambda t}\right|^2 = \frac{1}{2\pi T}\left[\left(\sum_{t=1}^{T} x_t \cos\lambda t\right)^2 + \left(\sum_{t=1}^{T} x_t \sin\lambda t\right)^2\right].$$

Evaluate $I(\lambda)$ at points $\lambda_j = 2\pi j/T$ for $j = 1, \ldots, [T/2]$.

Note that

$$\log(I(\lambda_j)) = c - 2d\log(\lambda_j) + \log(g(\lambda_j)/g(0)) + \varepsilon_j \qquad (1)$$

where $c = \log(g(0)) - C$ and $\varepsilon_j = \log[I(\lambda_j)/f(\lambda_j)] + C$

- C is Euler’s constant. 0.577216...
- It can be shown that ε_j are uncorrelated with mean zero in large samples.

Geweke and Porter Hudak 1983 (GPH) estimator

GPH suggested running the regression

$$\log(I(\lambda_j)) = c + dX_j + U_j$$

where $X_j = -2\log(\sin\lambda_j/2)$ for periodogram points $j = 1,\ldots,M$ for $M = O(T^{1/2})$.

$$\hat{d}_{GPH} = \frac{\sum_{j=1}^{M}(X_j - \bar{X})\log(I(\lambda_j))}{\sum_{j=1}^{M}(X_j - \bar{X})^2}$$

- GPH argued that for points in the neighbourhood of 0, $\log(g(\lambda_j)/g(0)) \approx 0$.
- Hence, the bias in $\hat{d}_{GPH}$ should be small.
- It can be shown (Hurvich, Deo and Brodsky 1998) that

$$E(\hat{d} - d) = O(M^2/T^2)$$

$$RMSE(\hat{d}) = O(M^2/T^2)$$

- Also, if $M = o(T^{4/5})$

$$M^{1/2}(\hat{d} - d) \underset{asy}{\sim} N(0,\pi^2/24).$$

There are several variants of the GPH method, such as

- Different truncation rates.
- Trimming periodogram points for j close to 0.
- Smoothing periodogram points.
- Setting $X_j = -2\log(\lambda_j/2)$ (equivalent to GPH in large samples).

Results for the Nile Data

GPH with $M = 90$: $d = 0.4059\ (0.0813)$

Moulines and Soulier 2000 (MS) "Broadband" Method

MS note the Fourier expansion

$$\log(g(\lambda)) = \sum_{j=1}^{\infty} \theta_j h_j(\lambda)$$

where

$$h_j(\lambda) = \cos(j\lambda)/\pi.$$

They propose the regression

$$\log(I(\lambda_j)) = c - 2d\log(\lambda_j) + \theta_1 h_1(\lambda_j) + \cdots + \theta_P h_P(\lambda_j) + U_j$$

- Set $M = [T/2]$.
- model $\log(g(\lambda_j))$ by including dummies $h_1(\lambda), \ldots h_P(\lambda_j)$ in the regression.
- Choose $P \to \infty$ but $P/T \to 0$ as $T \to \infty$.
- This estimator is also consistent and asymptotically normal.
- P can increase more slowly than T/M (M from GPH estimator) and this estimator can accordingly converge faster
- However, the dummies are highly collinear in finite samples beyond a certain point.

Results for Nile Data $(P = 4)$: $d = 0.4451\ (0.09)$

Fractional Integration

Now consider a AR(∞) process $x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots = u_t$ where for a parameter $d \leq 1$,

$$a_j = a_{j-1}\frac{j-d-1}{j}, \quad j \geq 1$$

Alternative formulae are

$$a_j = \begin{cases} \dfrac{-d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} & d \geq 0, \\[2ex] \dfrac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} & d < 0 \end{cases}$$

It is well known from Stirling's approximation that $a_j = O(j^{-d-1})$.

- $d = 0$: $a_j = 0$ for $j > 0$
- For $d = 1$: $a_1 = -1$, and $a_j = 0$ for $j > 1$.
- For $d = -1$, $a_j = 1$ for $j > 0$

More generally, for real values of d, the identity

$$\sum_{j=0}^{\infty} a_j L^j = (1 - L)^d$$

follows from the binomial expansion of $(1 - L)^d$.

This is the *fractional difference operator.*

The MA(∞) representation of the process is

$$x_t = (1 - L)^{-d} u_t = \sum_{j=0}^{\infty} b_j u_{t-j}$$

where for $d > 0$, $b_j = \dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sim \dfrac{j^{d-1}}{\Gamma(d)}$.

Properties of the Fractional Integration model.

Assume that $u_t \sim \text{i.i.d.}(0,\sigma^2)$.

Consider for $0 < d < \frac{1}{2}$,

$$x_t = (1-L)^{-d}u_t = \sum_{j=0}^{\infty} b_j u_{t-j}.$$

Then since $b_j^2 = O(j^{2(d-1)})$,

$$\text{Var}(x_t) = \sigma^2 \sum_{j=0}^{\infty} b_j^2 < \infty$$

This process is covariance stationary.

It can be shown that the same property holds when u_t is a weakly dependent process.

For $\frac{1}{2} \leq d < \frac{3}{2}$, consider the stationary process

$$\Delta x_t = (1 - L)^{1-d} u_t.$$

It can be shown that

$$E(x_T^2) = E(\Delta x_1 + \Delta x_2 + \cdots + \Delta x_T)^2 = O(T^{2d-1}).$$

Hence, x_t is not covariance stationary in this case.

However, in the case $d < 1$ it is independent of initial conditions, since then

$$b_j = O(j^{d-1}) \to 0$$

as $j \to \infty$.

The case $d = 1$ is the usual ‘integrated’ (I(1)) process.

Fractional integration, or I(d) therefore interpolates between I(0) and I(1).

Spectrum of the Stationary FI process

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda)$$

where $0 < f^*(0) < \infty$. Note that

$$|1 - e^{-i\lambda}|^2 = (1 - \cos\lambda - i\sin\lambda)(1 - \cos\lambda + i\sin\lambda)$$

$$= 2(1 - \cos\lambda) = \lambda^2 + O(\lambda^4)$$

Hence, $f(\lambda) = |\lambda|^{-2d} g(\lambda)$ where

$$g(\lambda) = \frac{f^*(\lambda)}{(1 + O(\lambda^2))^d}$$

Example: the ARFIMA(p, d, q) Model

In the time domain, this has the form

$$\phi(L)(1 - L)^d x_t = \theta(L) u_t$$

where

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$$

The spectral density of the process is

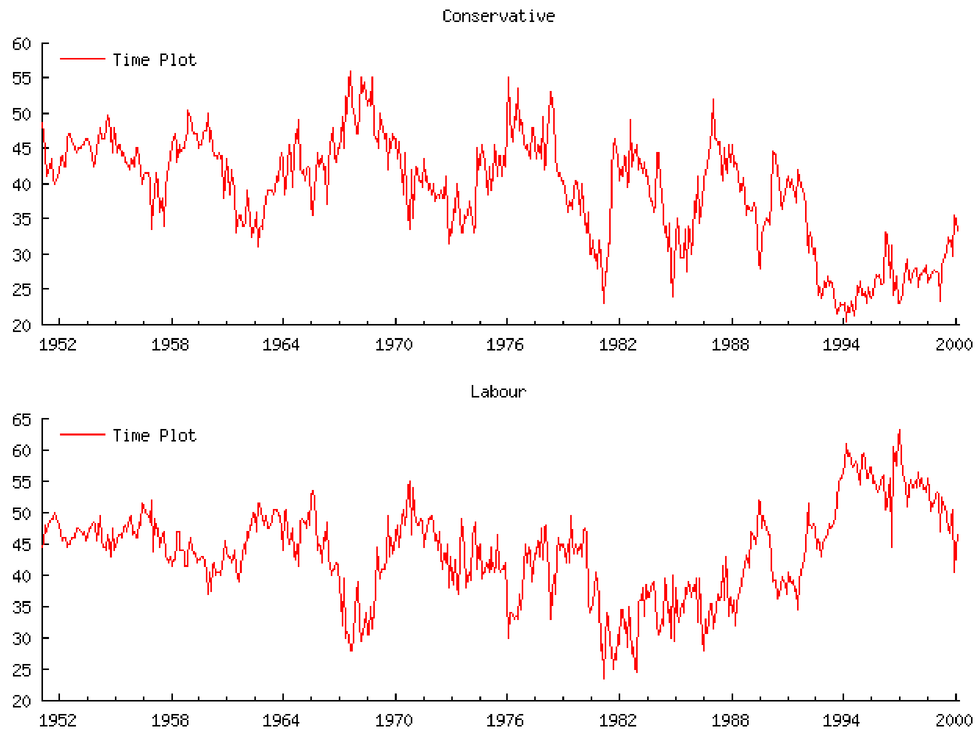
$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$$

Estimating the ARFIMA model

- Time domain: Least squares, conditional Maximum Likelihood.
- Frequency domain: Whittle ML, based on periodogram points.

Nonstationary Long Memory

Example: opinion poll data



If $d > 0.5$, one can estimate $d - 1$ from the differenced data. GPH yields (after adding back 1):

Conservative: GPH $d = 0.7744(0.0875)$

Labour:: GPH $d = 0.7685\ (0.0875)$

Asymptotic Theory for Long Memory Models.

Consider a stationary fractionally integrated process

$$x_t = (1-L)^{-d}u_t = \sum\nolimits_{j=0}^{\infty} b_j u_{t-j}, \quad |d| < \tfrac{1}{2}.$$

$$X_T(r) = \frac{1}{\sigma_x\sqrt{T}} \sum\nolimits_{t=1}^{[Tr]} x_t$$

cannot converge to BM, since the conditions for the FCLT are violated.

However, note the partial sum representation

$$X_T(r) = \frac{1}{\sigma_u T^{d+1/2}} \sum\nolimits_{t=1}^{[Tr]} \sum\nolimits_{j=0}^{\infty} b_j u_{t-j} = \frac{1}{\sigma_u T^{d+1/2}} \sum\nolimits_{k=-\infty}^{[Tr]} a_{Tk}(r) u_k$$

where

$$a_{Tk}(r) = \sum\nolimits_{j=0}^{[Tr]-k} b_j, \quad k \ge 0, \qquad a_{Tk}(r) = \sum\nolimits_{j=-k}^{[Tr]-k} b_j, \quad k < 0.$$

An FCLT can be proved for these sums.

The limit process is *fractional Brownian motion* (fBM):

$$B_d(r) = \frac{1}{\Gamma(1+d)}\left(\int_0^r (r-s)^d dB + \int_{-\infty}^0 [(r-s)^d - (-s)^d]dB\right),\ \ 0 \le r \le 1.$$

Points

- Note modified normalization, variance of sum = $O(T^{2d+1})$.
- Formula reduces to $B(r)$ in the case $d = 0$.
- The continuous mapping theorem can be invoked as before, to derive limits as functionals of fBM.
- The problem of fractional stochastic integrals requires a new approach, since the limit random variables are not Itô integrals.
- See Davidson and Hashimzade (ET 2008, 2009) for results in the frequency domain and time domain respectively.

A Rival Approach

Some authors work with so-called "Type 2" fBM, with the form

$$B_d^*(r) = \frac{1}{\Gamma(1+d)} \int_0^r (r-s)^d dB \qquad 0 \le r \le 1.$$

- This makes some of the asymptotic analysis easier, although the implied limit distributions are different to the "Type I" case above.
- However, it has slightly bizarre modelling implications. Letting $u_t^* = u_t I(t \ge 1)$, one must write the fractional increments in the form

$$x_t = (1-L)^{-d} u_t^* = \sum_{j=0}^{t-1} b_j u_{t-j}.$$

- If this is the correct model of $x_1, \ldots x_T$, it cannot be the correct model of $x_{1+K}, \ldots x_{T+K}$ for any $K \neq 0$. (?!)

Origins of Long Memory

Long memory poses a problem for economic modelling.

Most dynamic economic models are constructed as difference equations (linear or nonlinear).

For example

$$x_t = f(v_t, x_{t-1})$$

For stability $f(.,v_t)$ must be a *contraction* mapping, with the property

$$\sup_{x,v} \left| \frac{\partial f(v,x)}{\partial x} \right| \leq b < 1.$$

Imagine solving this for $J \geq 1$ as

$$x_t = f(v_t, f(v_{t-1}, f(v_{t-2} \cdots f(v_{j-J}, x_{t-J-1}) \cdots)) = g(v_t, v_{t-1}, v_{t-2} \cdots, v_{j-J+1}, x_{t-J})$$

Note that

$$\left| \frac{\partial g}{\partial v_{t-J}} \right| \leq b^{J-1} \left| \frac{\partial f}{\partial v_{t-J}} \right|.$$

This is *exponential* memory decay (i.e. short memory).

For the case $b = 1$, on the other hand, nonstationary behaviour (unit roots) is generated.

This kind of model cannot produce long memory.!

Granger Aggregation (Granger JoE 1980)

Most useful models of long memory generation are based on some notion of *aggregation of micro-processes.*

Suppose

$$x_{jt} = \alpha_j x_{j,t-1} + \beta_j W_t, \quad j = 1,\dots,N$$

where $\alpha_1,\dots\alpha_N$ and $\beta_1,\dots,\beta_N$ are independent drawings from given distributions with $\alpha_j \in [0,1]$ w.p.1.

Then,

$$\bar{x}_t \overset{def}{=} \frac{1}{N}\sum_{j=1}^{N} x_{jt} = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\beta_j}{1-\alpha_j L}\right)W_t$$

where

$$\frac{1}{N}\sum\nolimits_{j=1}^{N}\left(\frac{\beta_j}{1-\alpha_j L}\right) \simeq E(\beta)\int \frac{1}{1-\alpha L}dF(\alpha)$$

If $F(\alpha)$ represents the Beta(p,q) distribution, then

$$dF(\alpha) = \frac{1}{B(p,q)}\alpha^{2p-1}(1-\alpha^2)^{q-1}d\alpha, \quad 0 \leq \alpha \leq 1$$

where $B(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ and

$$\bar{x}_t \simeq E(\beta)\sum_{k=0}^{\infty}\frac{B(p+k/2,q)}{B(p,q)}W_{t-k} = E(\beta)\sum_{k=0}^{\infty}\frac{\Gamma(p+k/2)\Gamma(p+q)}{\Gamma(p+k/2+q)\Gamma(p)}W_{t-k}$$

where

$$\frac{B(p+k/2,q)}{B(p,q)} = O(k^{-q}).$$

Setting $d = 1-q$, note that $\bar{x}_t \sim I(d)$.

- Easily generalized with (e.g.) idiosyncratic shocks
- The model is linear, and hence can show fBM as the limit process, when $d > 0.5$.

Regime-Switching (Davidson and Sibbertsen JoE 2005)

Let

$$x_t = m_t + \varepsilon_t$$

where $m_t = k_j$ for $S_{j-1} < t \le S_j$, $j = 1,2,,\ldots$ where $\tau_j = S_j - S_{j-1}$ and

1. $\{\tau_j\}$ is distributed according to a power law, with

$$P(\tau = c) \sim c^{-1-\alpha}L(c), \quad 0 < \alpha < 2$$

2. $\{k_j\}$ is a short-memory process in "regime time".
3. $\{\varepsilon_t\}$ is a short-memory process in "calendar time".

Think of x_t as random variations round a local mean.

- The power law in "regime time" implies "bunching" of regime switches.
- Long regimes have relatively low probability, but account for a significant amount of calendar time.

Cases:

1. $1 < \alpha < 2$: Process $\{x_t\}$ is stationary, autocovariances exhibit long memory. with $H = (3 - \alpha)/2$
2. $0 < \alpha < 1$: Process $\{x_t\}$ is nonstationary, Δx_t is anti-persistent $H = (1 - \alpha)/2$.

Long Memory vs Fractional Integration

These regime-switching processes are long memory, but are nonlinear.
They are *not* fractionally intregrated processes, and do *not* converge to fBM.

- In Case 1, time aggregation ($T \to \infty$) can be shown to yield an I(1) process with infinite-variance increments ("Levy motion").

- However, cross-sectional aggregation of independent micro-processes (or dependent Gaussian micro-processes) *can* yield fBM, similarly to Granger case.

- Important to distinguish general long memory from FI processes.

- Only processes with linear representation in the limit allow inference based on fBM.