Lecture 5:
Mean-Reversion

Marco Avellaneda

G63.2936.001

Spring Semester 2009

Stationarity/ Non Stationarity

Definition: a stochastic process is stationary if

$$\forall m, \forall (t_1, \ldots, t_m), \forall A \in \mathbf{R}^n$$
$$\Pr.\{(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \in A\} = \Pr.\{(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_m+h}) \in A\}$$

A stationary process is a process that is ``statistically invariant under translations''

Examples: the Ornstein-Uhlembeck process is stationary, Brownian motion is not.

The Ornstein-Uhlenbeck process

$$dX_t = \kappa(m - X_t)dt + \sigma dW_t$$

$$X_t = e^{-\kappa(t-s)}X_s + \left(1 - e^{-\kappa(t-s)}\right)m + \sigma\int_s^t e^{-\kappa(t-u)}dW_u$$

$$X_t = m + \sigma\int_{-\infty}^t e^{-\kappa(t-s)}\eta(s)ds, \quad \eta(s) = \text{Gaussian white noise}$$

Exponentially-weighted moving average of uncorrelated Gaussian random variables.

Statistics of the OU process

$$\langle X_t X_{t+h} \rangle = \sigma^2 \left\langle \int_{-\infty}^{t} e^{-k(t-s)} \eta(s) ds \cdot \int_{-\infty}^{t+h} e^{-k(t+h-s')} \eta(s') ds' \right\rangle$$

$$= \sigma^2 \int_{-\infty}^{t} \int_{-\infty}^{t+h} e^{-k(t-s)} e^{-k(t+h-s')} \delta(s-s') ds ds'$$

$$= \sigma^2 \int_{-\infty}^{t} e^{-k(t-s)} e^{-k(t+h-s)} ds$$

$$= \sigma^2 e^{-kh} \int_{-\infty}^{t} e^{-2k(t-s)} ds$$

$$= \frac{\sigma^2 e^{-kh}}{2k}$$

$$\left\langle \left| X_{t+h} - X_t \right|^2 \right\rangle = \frac{\sigma^2}{k} \left(1 - e^{-kh} \right)$$

Structure Function

Random Walk, Fractional BM

$$X_t = \sigma W_t, \qquad W_t = \text{Brownian motion}$$

$$\left\langle \left| X_{t+h} - X_t \right|^2 \right\rangle = \sigma^2 h \qquad \left\langle X_{t+h} X_t \right\rangle = t$$

Brownian motion (non-stationary)
Structure fn grows linearly

$$X_t = \sigma \int_{-\infty}^{t} \frac{\eta(s)ds}{(1+t-s)^p} \qquad p > 1/2$$

Geometric Brownian motion

$$\left\langle X_t X_{t+h} \right\rangle = \frac{\sigma^2}{h^{2p-1}} \int_{\frac{1}{h}}^{\infty} \frac{du}{u^p (1+u)^p}$$

$$\left\langle X_t X_{t+h} \right\rangle \approx \begin{cases} \dfrac{\sigma^2}{h^{2p-1}} & 1/2 < p < 1 \\ \dfrac{\sigma^2 \ln(h)}{h} & p = 1 \\ \dfrac{\sigma^2}{h^p} & p > 1 \end{cases}$$

Correlations decay like power-laws
(large *h*)

Autoregressive Models

$$X_1, X_2, \ldots, X_n, \ldots$$

$$X_{n+1} = a_0 + a_1 X_n + \ldots + a_m X_{n-m+1} + \sigma \nu_{n+1}, \quad \nu_i \sim N(0,1)$$

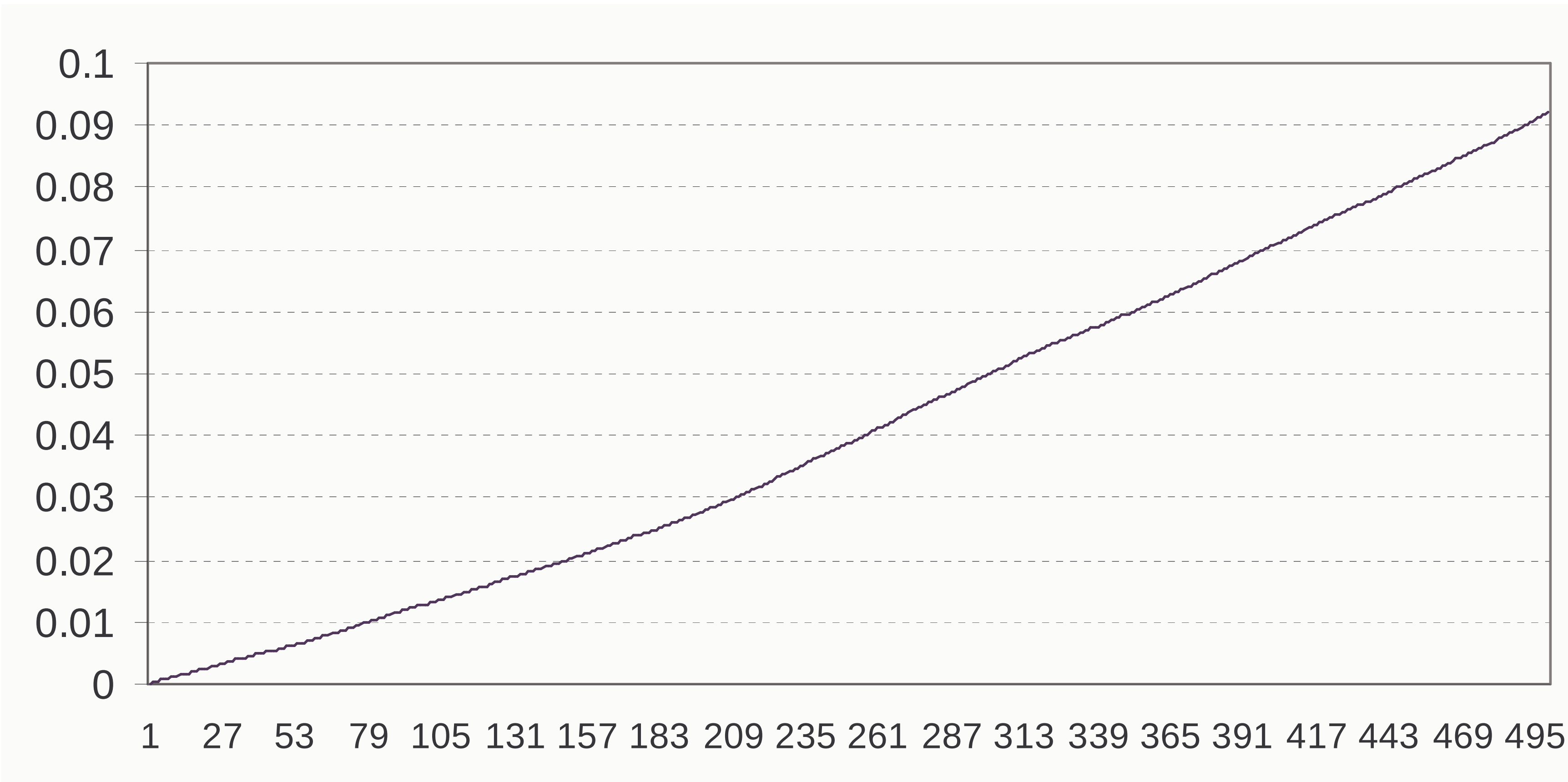
$$\mathbf{Y}_n = \begin{pmatrix} X_n \\ \ldots \\ X_{n-m+1} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_1 & \ldots & a_m \\ 1 & 0 & \ldots \\ \ldots & 1 & 0 \end{pmatrix}, \quad \mathbf{A}_0 = \begin{pmatrix} a_0 \\ 0 \\ \ldots \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma \\ 0 \\ \ldots \end{pmatrix}$$

$$\mathbf{Y}_{n+1} = \mathbf{A}_0 + \mathbf{A}\mathbf{Y}_n + \mathbf{\Sigma} \nu_{n+1}$$

AR-n model corresponds to a vector AR-1 model

Structure function: SPY Jan 1996-Jan 2009

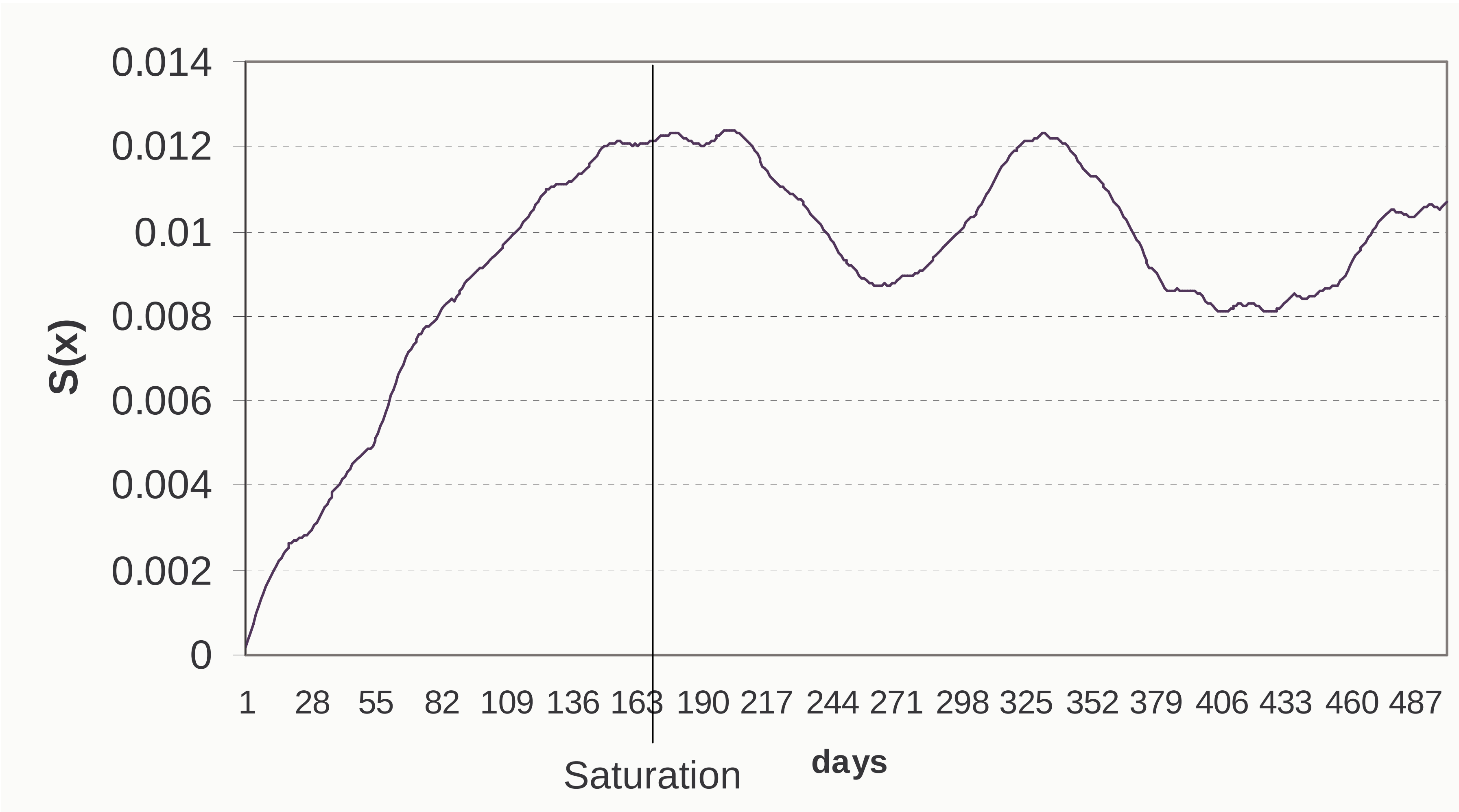
Use log prices as time series. Structure function with lags 1 day to 2 yrs



SPY is highly non stationary, as shown in the chart.

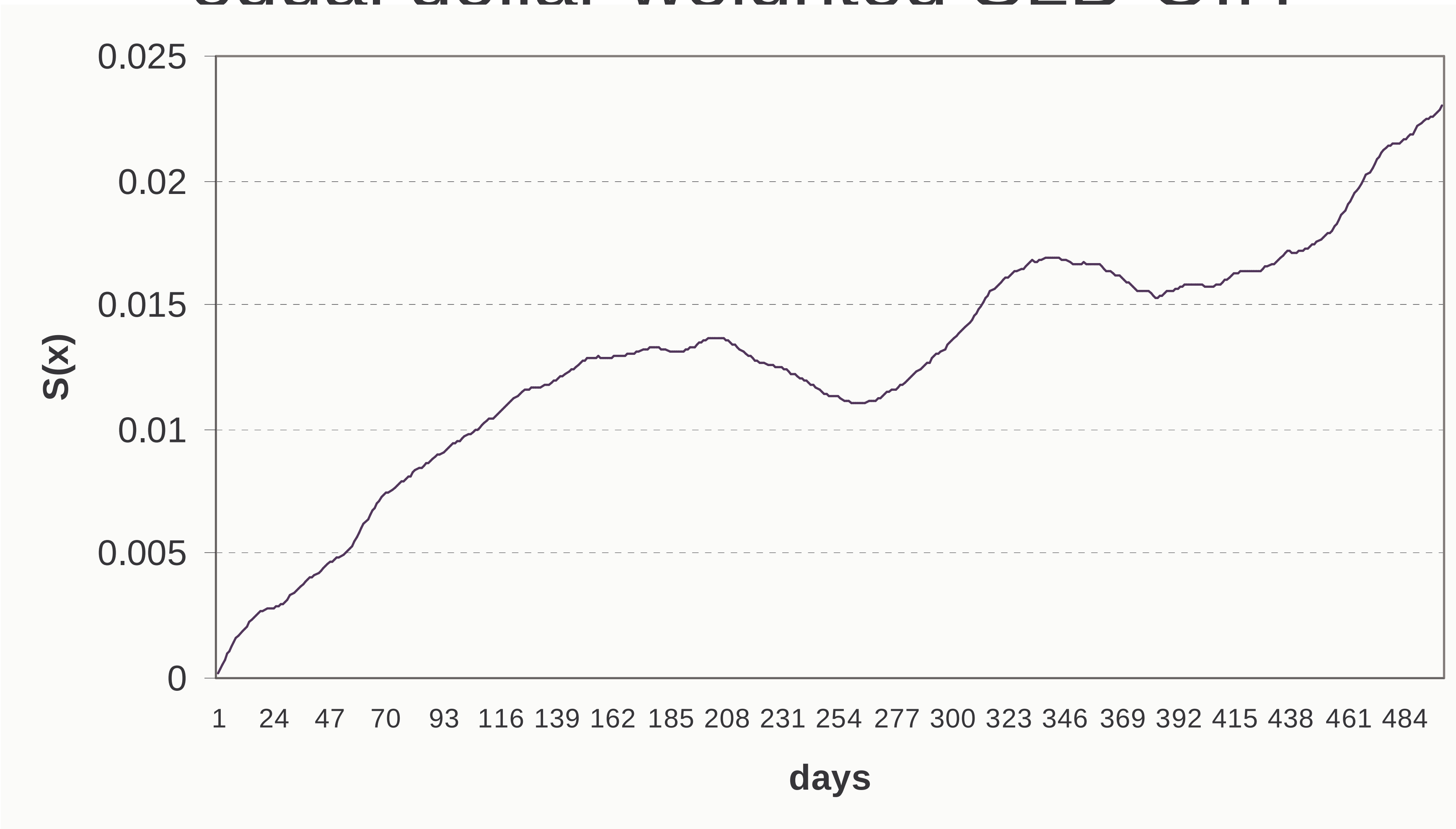
Look for mean-reversion in relative value, i.e. in terms of two or more assets.

Structure function log (SLB/OIH)
Data: Apr 2006 to Feb 2009



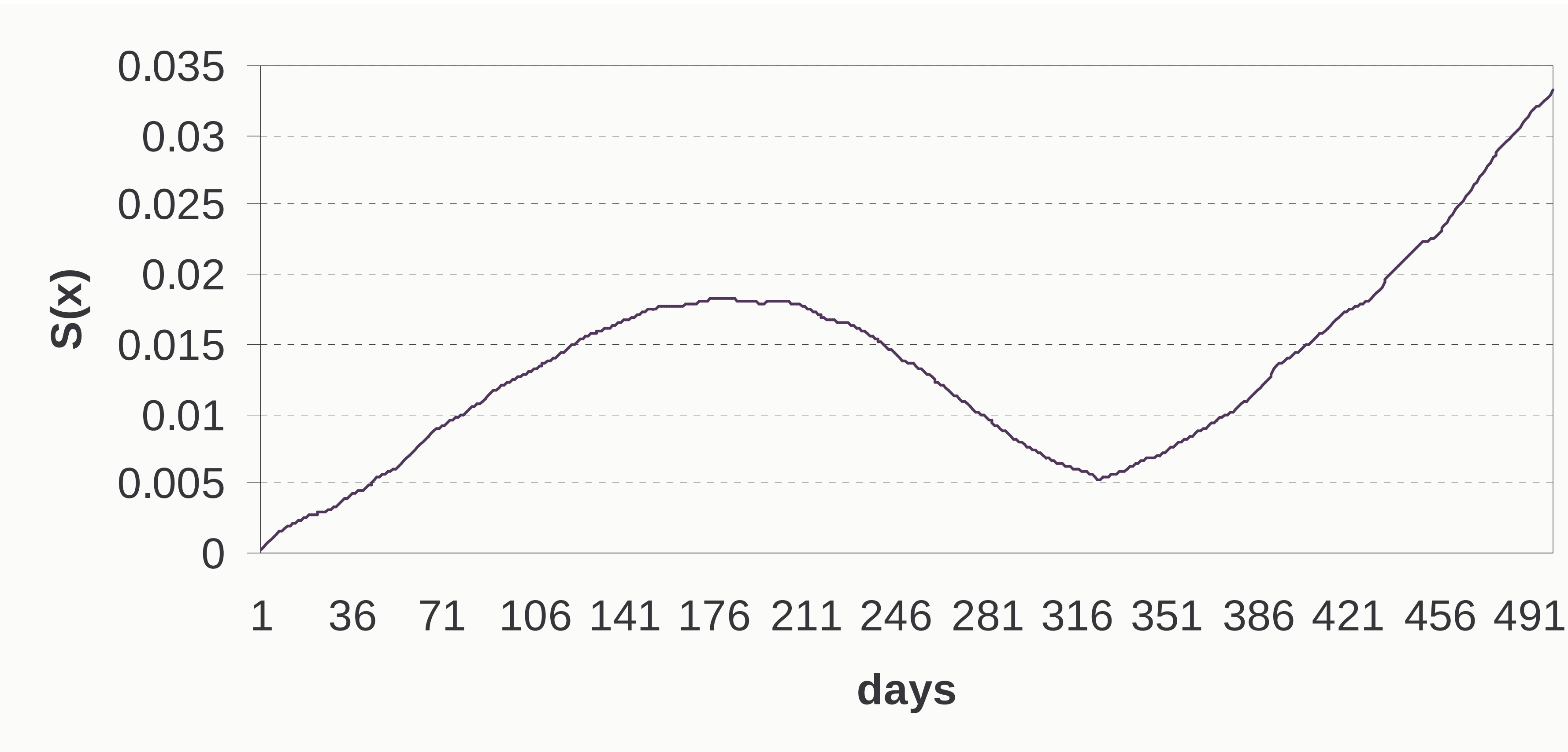
OIH: Oil Services ETF, SLB: Schlumberger-Doll Research

Structure Function: long-short equal dollar weighted SLB-OIH



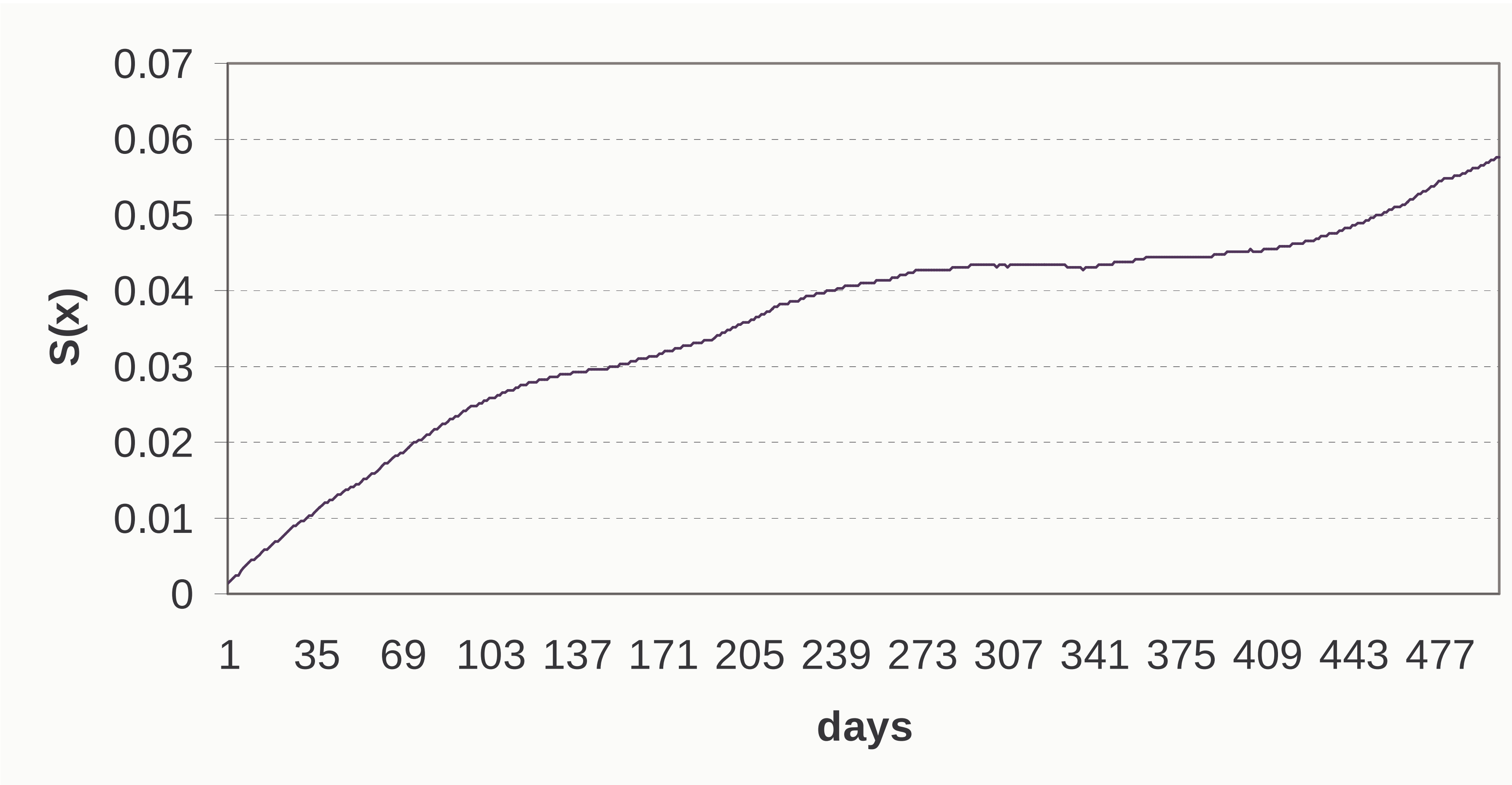
$$P_{n+1} = P_n \times \left(1 + R_{\mathrm{slb}} - R_{\mathrm{oih}}\right), \quad X_n = \ln P_n$$

Structure Function for Beta-Neutral long-short portfolio SLB-Beta*OIH

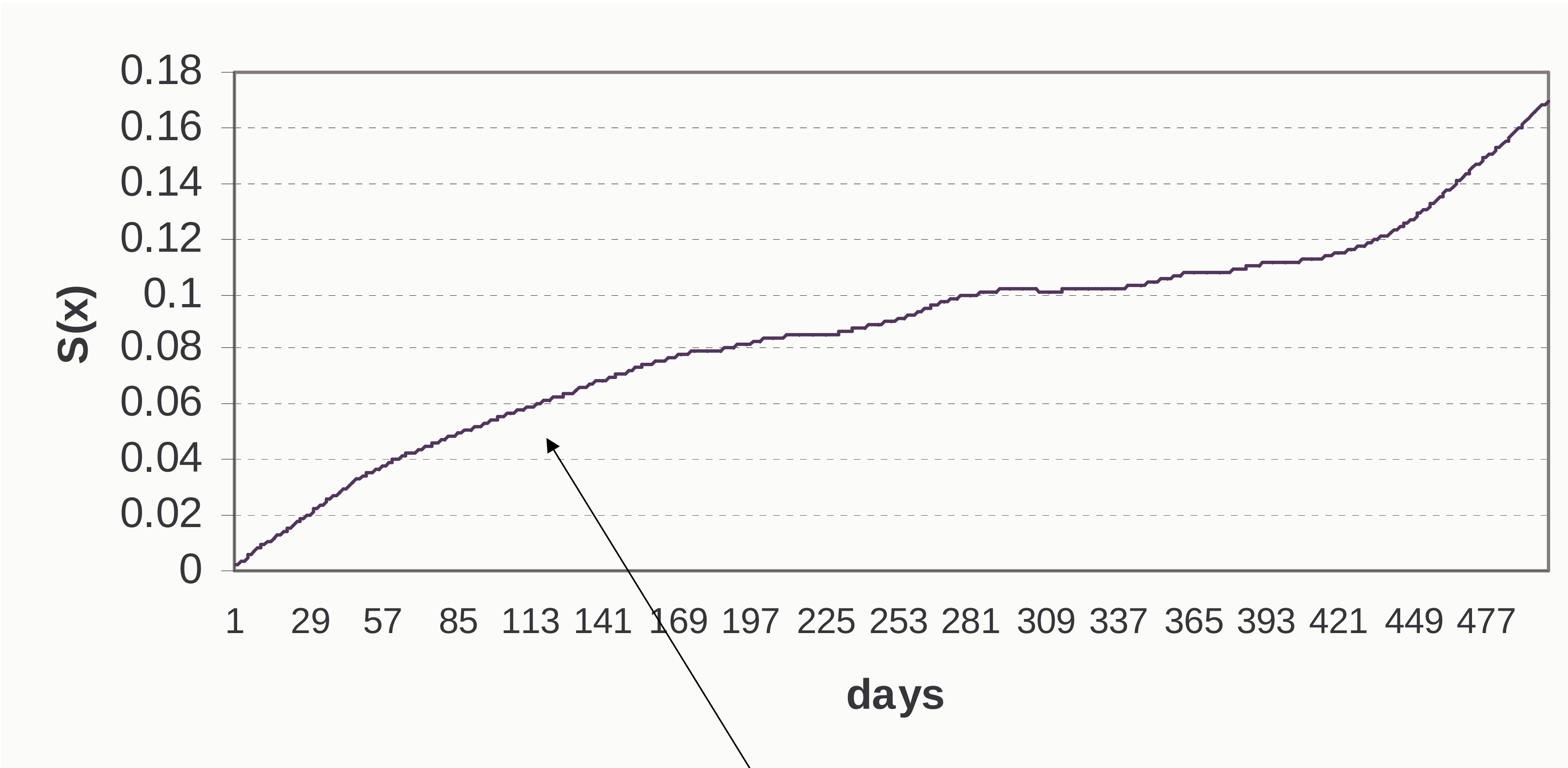


$$P_{n+1} = P_n \times \left(1 + R_{\mathrm{slb}} - \beta_{60\mathrm{d}} \cdot R_{\mathrm{oih}}\right), \quad X_n = \ln P_n$$

Structure Function log (GENZ/IBB)



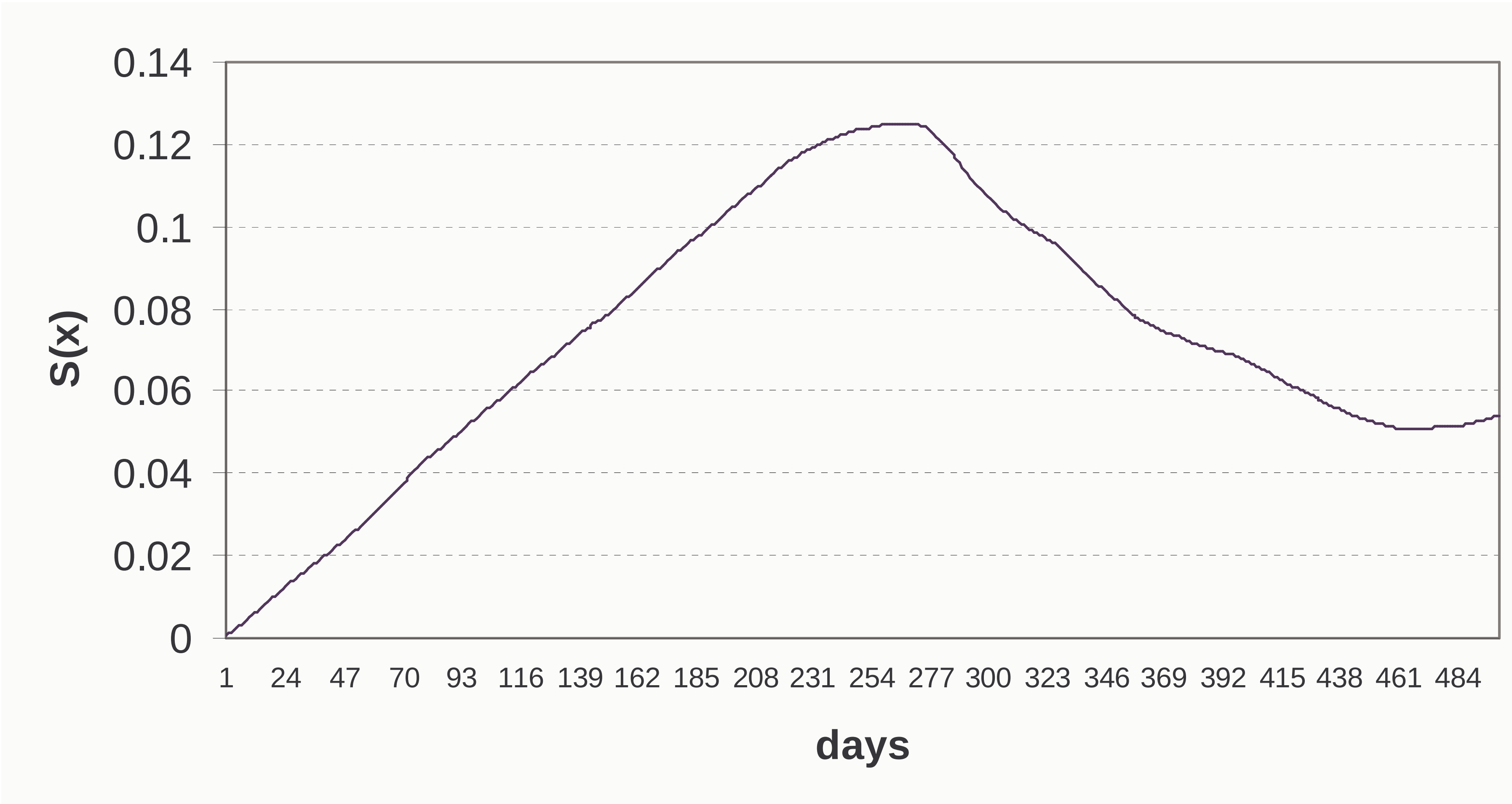
Structure function ln (DNA/GENZ)



DNA: Genentech Inc.
GENZ; Genzyme Corp.

Mean-reversion: large negative curvature here.

Structure Fn for Beta-Neutral GENZ-DNA Spread



Poor reversion for the beta adjusted pair. Beta is low ~ 0.30

Systematic Approach for looking for MR in Equities

Look for stock returns devoid of explanatory factors, and analyze the corresponding residuals as stochastic processes.

$$R_t = \sum_{k=1}^{m} \beta_k F_{kt} + \varepsilon_t$$

Econometric factor model

$$X_t = X_0 + \sum_{s=1}^{t} \varepsilon_s$$

View residuals as increments of a process that will be estimated

$$\frac{dS(t)}{S(t)} = \sum_{k=1}^{m} \beta_k \frac{dP_k(t)}{P_k(t)} + dX(t)$$

Continuous-time model for evolution of stock price

Interpretation of the model

The factors are either

A. eigenportfolios corresponding to significant eigenvalues of the market

B. industry ETF, or portfolios of ETFs

Questions of interest:

Can residuals be fitted to (increments of) OU processes or other MR processes?

If so, what is the typical correlation time-scale?

Estimation of Ornstein-Uhlenbeck models

$$X_{t+\Delta t} = e^{-k\Delta t} X_t + m\left(1 - e^{-k\Delta t}\right) + \sigma \int_t^{t+\Delta t} e^{-k(t-s)} dW_s$$

$$X_{n+1} = aX_n + b + \nu_{n+1} \qquad \{\nu_n\} \text{ i.i.d. } N\left(0, \sigma^2\left(\frac{1-e^{-2k\Delta t}}{2k}\right)\right)$$

$$a = \text{SLOPE}\left(\left(X_{n-l},\ldots,X_n\right);\left(X_{n-l-1},\ldots,X_{n-1}\right)\right),$$

$$b = \text{INTERCEPT}\left(\left(X_{n-l},\ldots,X_n\right);\left(X_{n-l-1},\ldots,X_{n-1}\right)\right)$$

$$k = \frac{1}{\Delta t}\ln\left(\frac{1}{a}\right), \quad m = \frac{b}{1-a}, \quad \sigma = \frac{\text{STDEV}\left(X_{n+1} - aX_n - b\right)}{\sqrt{1-a^2}}\sqrt{2\frac{1}{\Delta t}\ln\left(\frac{1}{a}\right)}$$

Portfolio Strategy

$Q_1, Q_2, ..., Q_N$ $\$$ invested in different stocks (long or short)

$S_1, S_2, ..., S_N$ dividend-adjusted prices

$$d\Pi = \sum_{i=1}^{N} Q_i \frac{dS_i}{S_i} - \left(\sum_{i=1}^{N} Q_i \right) rdt \qquad \text{(neglect transaction costs)}$$

$$= \sum_{i=1}^{N} Q_i \left(\sum_{k=1}^{m} \beta_{ik} \frac{dP_k}{P_k} + dX_i \right) - \left(\sum_{i=1}^{N} Q_i \right) rdt$$

$$= \sum_{i=1}^{N} Q_i dX_i + \sum_{k=1}^{m} \left(\sum_{i=1}^{N} Q_i \beta_{ik} \right) \frac{dP_k}{P_k} - \left(\sum_{i=1}^{N} Q_i \right) rdt$$

$\sum_{i=1}^{N} Q_i \beta_{ik}$: net dollar-beta exposure along factor k

$\left(\sum_{i=1}^{N} Q_i \right)$: net dollar exposure of portfolio

Market-Neutral Portfolio

Assume $dX_i = k_i(m - X_i)dt + \sigma_i dW_i$ $\{dW_i\}_{i=1}^{N}$ uncorrelated

$$d\Pi = \sum_{i=1}^{N} Q_i dX_i - \left(\sum_{i=1}^{N} Q_i\right) r dt$$

$$= \sum_{i=1}^{N} Q_i \left(k_i(m - X_i)dt + \sigma_i dW_i\right) - \left(\sum_{i=1}^{N} Q_i\right) r dt$$

$$= \sum_{i=1}^{N} Q_i \left(k_i(m - X_i) - r\right)dt + \sum_{i=1}^{N} Q_i \sigma_i dW_i$$

$$\therefore$$

$$E(d\Pi \mid \mathbf{X}) = \sum_{i=1}^{N} Q_i \left(k_i(m - X_i) - r\right)dt$$

$$Var(d\Pi \mid \mathbf{X}) = \sum_{i=1}^{N} Q^2{}_i \sigma^2{}_i dt$$

Mean-Variance Optimal Portfolio

$$\max_Q \left(\sum_i Q_i \mu_i - \frac{1}{2\lambda} \sum_i Q_i^{\,2} \sigma_i^2 \right) \quad \therefore \quad Q_i = \lambda \frac{\mu_i}{\sigma_i^2}$$

$$(\text{if } \ r = 0, \ \text{ or} \sum Q_i = 0)$$

$$d\Pi = \lambda \sum_i \frac{k_i^2 (m - X_i)^2}{\sigma_i^2} dt + \lambda \sum_i \frac{k_i (m - X_i)}{\sigma_i} dW_i$$

$$d\Pi = \lambda \sum_i \frac{k_i}{2} \xi_i^2 dt + \lambda \sum_i \sqrt{\frac{k_i}{2}} \xi_i \, dW_i \qquad \xi_i = \frac{m - X_i}{\sigma_i} \sqrt{2k_i}$$

$$\langle d\Pi \rangle = \frac{\lambda N}{2} \left(\frac{\sum_i k_i}{N} \right) dt \quad ; \quad \langle (d\Pi)^2 \rangle - \langle d\Pi \rangle^2 = \frac{\lambda^2 N}{2} \left(\frac{\sum_i k_i}{N} \right) dt$$

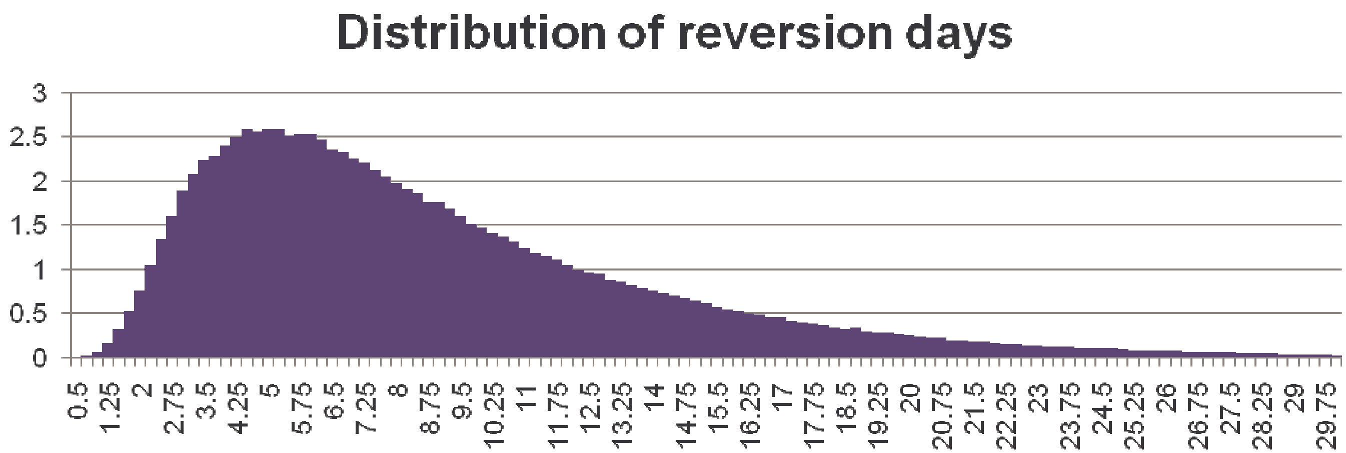
$$\text{Annualized Sharpe Ratio} = \sqrt{\frac{N}{2} \cdot \left(\frac{\sum_i k_i}{N} \right)} = \sqrt{\frac{N \bar{k}}{2}}$$

Statistics on the Estimated OU Parameters

ETF	Abs(Alpha)	Beta	Kappa	Reversion days	EquiVol	Abs(m)
HHH	0.20%	0.69	38	7	4%	3.3%
IYR	0.11%	0.90	39	6	2%	1.8%
IYT	0.18%	0.97	41	6	4%	3.0%
RKH	0.10%	0.98	39	6	2%	1.7%
RTH	0.17%	1.02	39	6	3%	2.7%
SMH	0.19%	1.01	40	6	4%	3.2%
UTH	0.09%	0.81	42	6	2%	1.4%
XLF	0.11%	0.83	42	6	2%	1.8%
XLI	0.15%	1.15	42	6	3%	2.4%
XLK	0.17%	1.03	42	6	3%	2.7%
XLP	0.12%	1.01	42	6	2%	2.0%
XLV	0.14%	1.05	38	7	3%	2.5%
XLY	0.16%	1.03	39	6	3%	2.5%
Total	0.15%	0.96	40	6	3%	2.4%

Average over 2006-2007

Mean reversion days: how long does it take to converge?



T_{days}=252/*k*

	Days
Max	30
75 %	11.4
Median	7.5
25 %	4.9
Min	0.5
Fast days	36%

Fast days : Percentage of faster mean reversion than 7 days