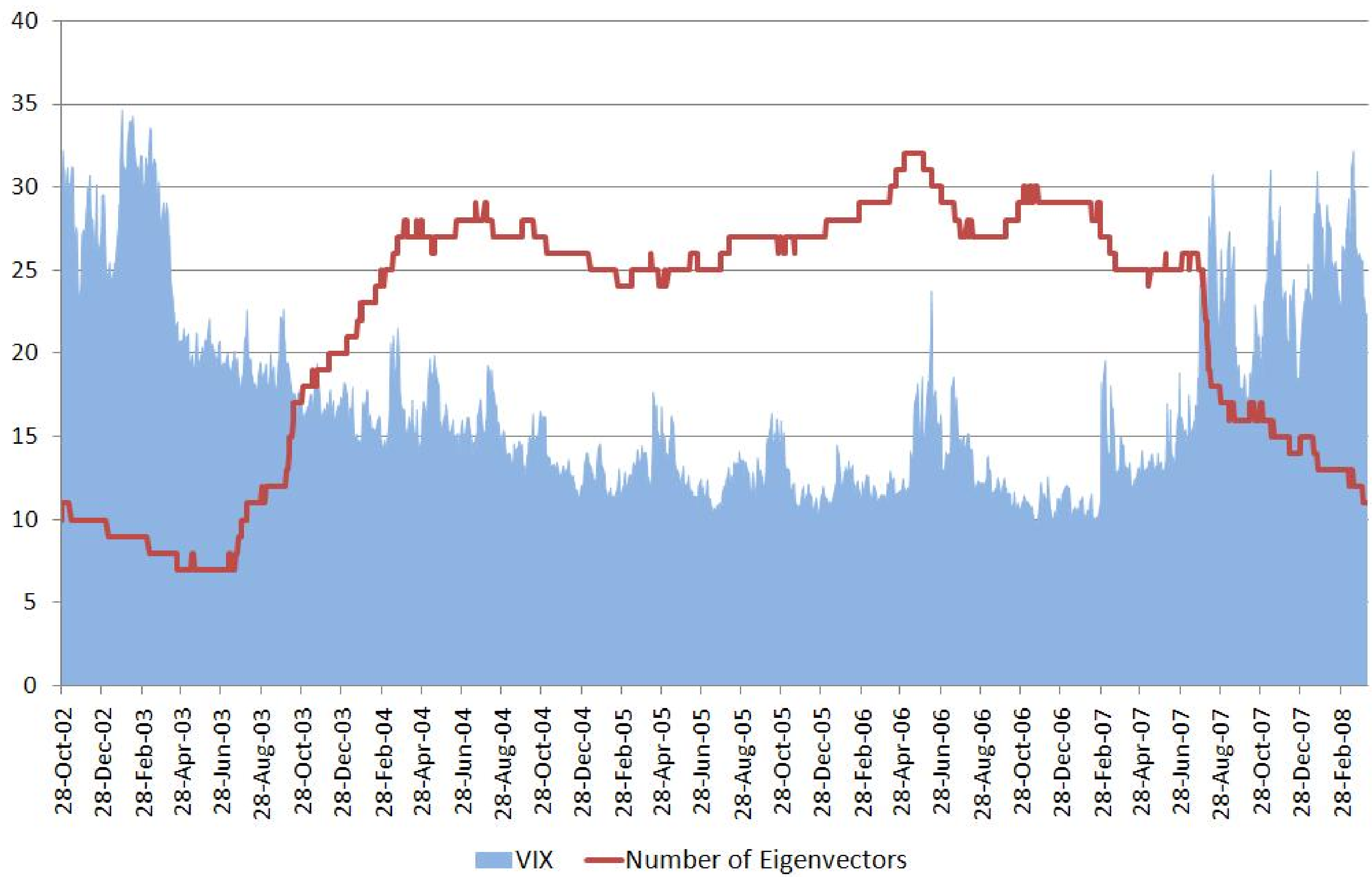
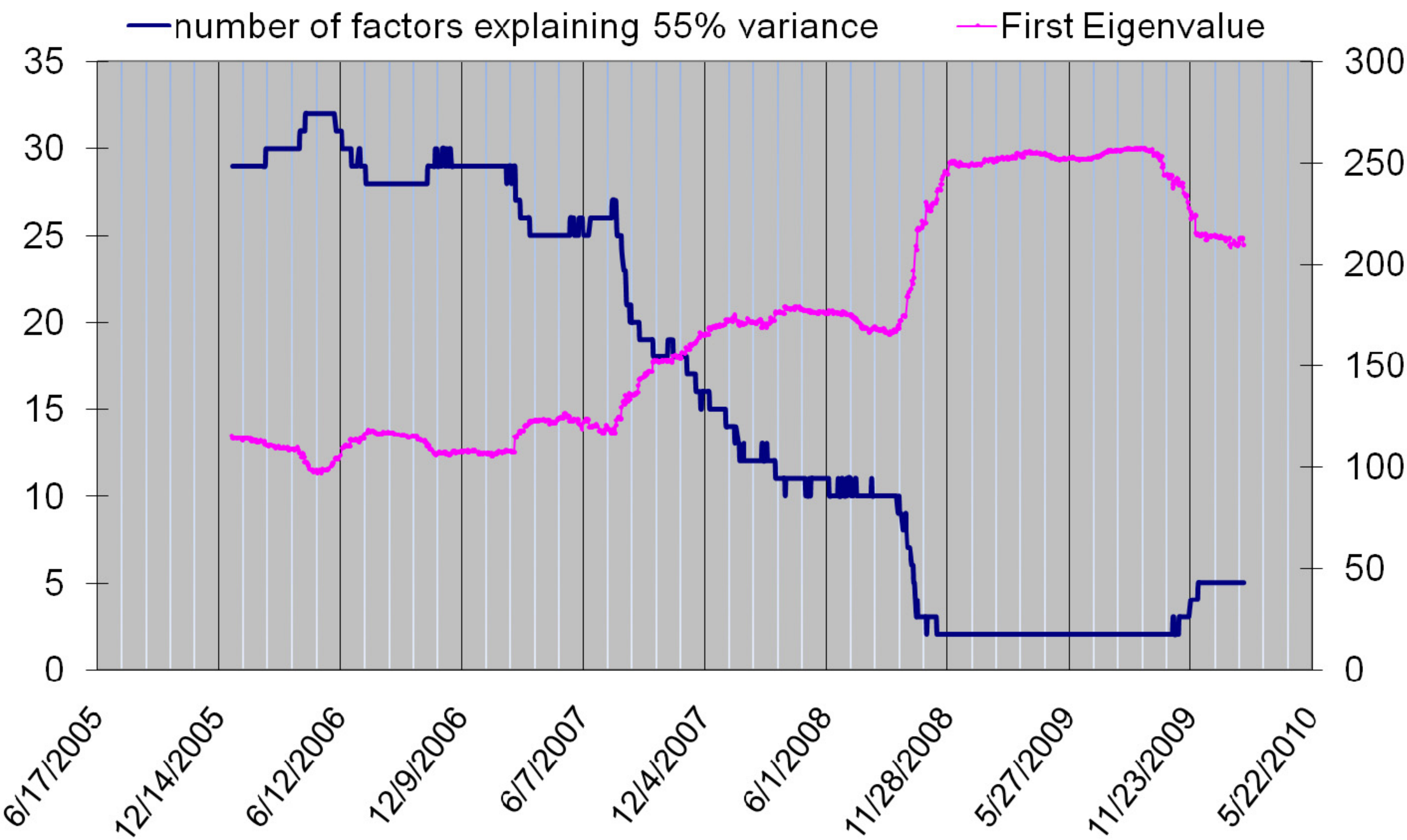
Risk and Portfolio Management
Spring 2010

Stochastic Processes & dynamics of stock prices

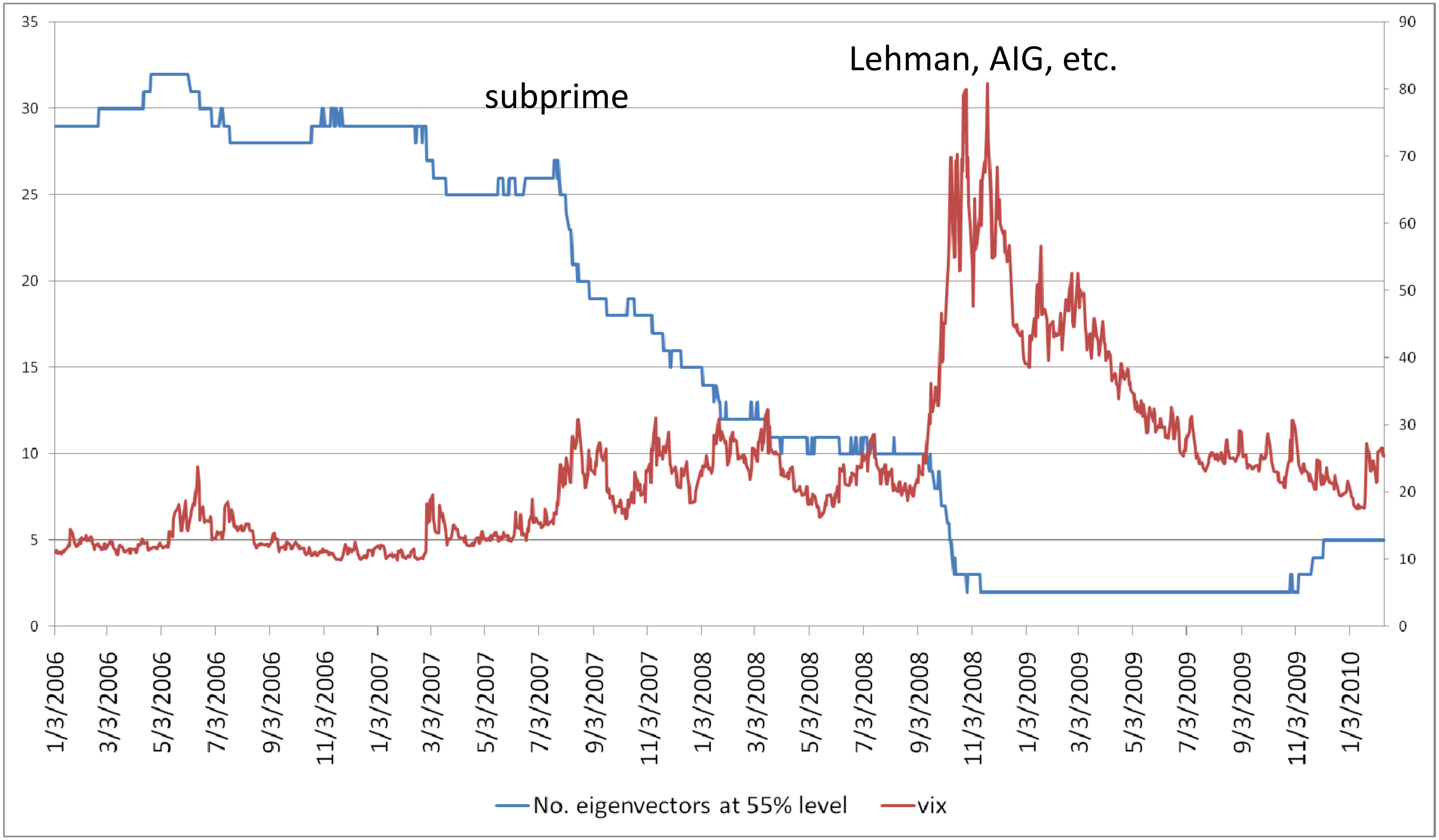
Number of factors explaining 55% of the variance versus VIX volatility index (2002-2008)



Number of explanatory factors vs. first eigenvalue of correlation matrix



Number of EVs versus VIX (1/2006-2/2010)



Dynamics are important

The previous slides show that the structure of the market is far from static.

This is obvious if we consider innovations in the market (new issues, new industries, the economic cycle, bubbles).

Equilibrium theories (e.g. APT, CAPM) are insufficient to explain prices, volatilities and correlations of financial assets.

Hence the need to model the evolution of financial variables using stochastic processes based on time-series analysis.

What can time-series analysis do for us?

-- Understand serial correlations in the data
-- Construct predictive models over suitable time-windows.
-- Discrete-time processes: important for data analysis.
-- Continuous-time processes: useful for theoretical purposes and to model high-dimensional data.

Stationarity/ Non Stationarity

Definition: a stochastic process is stationary if

$$\forall m, \quad \forall(t_1,\ldots,t_m), \qquad \forall E \in \mathbf{R}^n$$

$$\Pr.\{(X_{t_1}, X_{t_2},\ldots, X_{t_m}) \in A\} = \Pr.\{(X_{t_1+h}, X_{t_2+h},\ldots, X_{t_m+h}) \in A\}$$

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

$$dX_t = \kappa(m - X_t)dt + \sigma dW_t, \quad \kappa > 0$$

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

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

Exponentially-weighted moving average of uncorrelated Gaussian random variables.

Serial correlations of the OU process

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

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

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

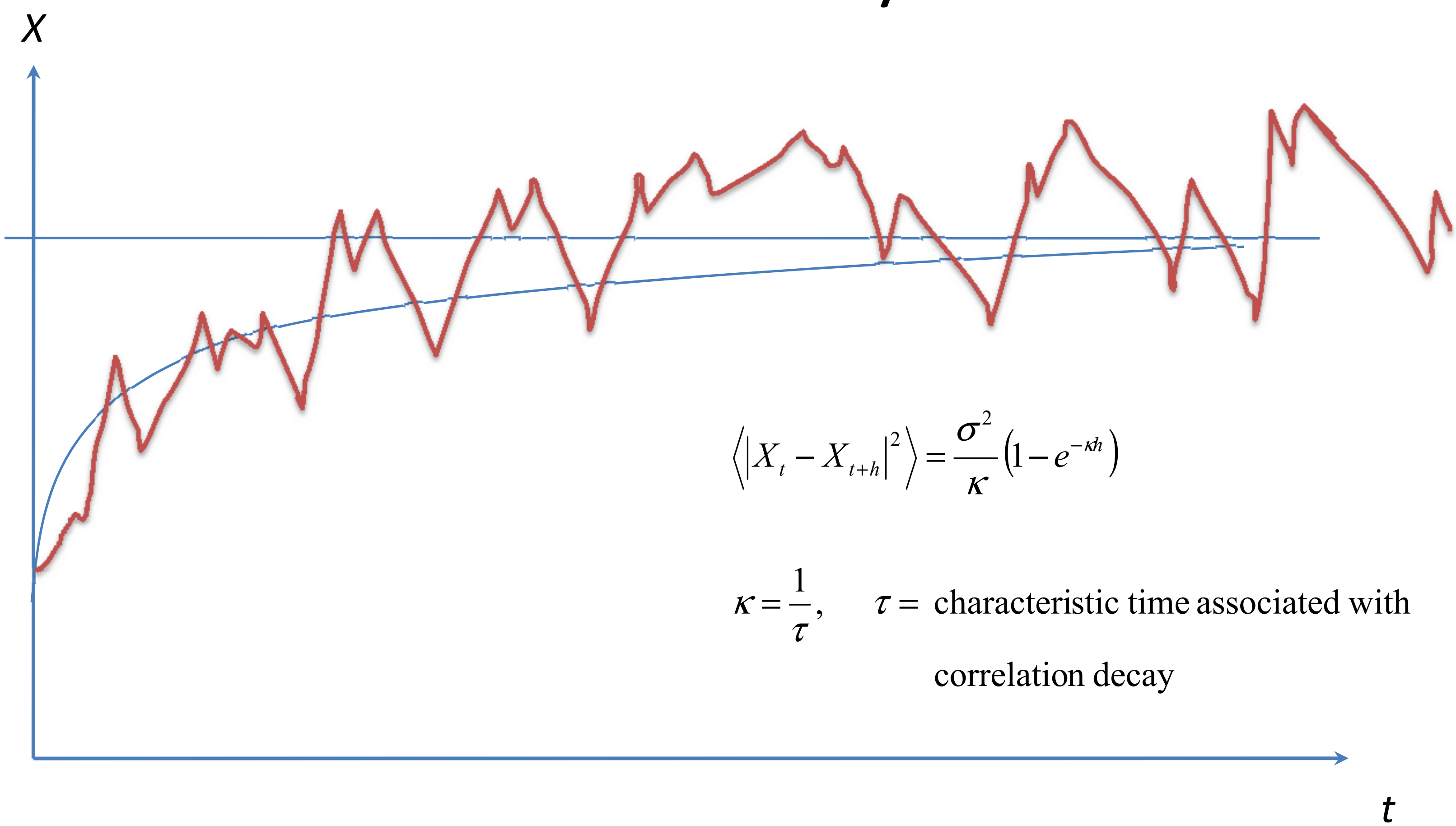
$$= \sigma^2 e^{-kh} \int_{-\infty}^{t} e^{-2k(t-s)} ds$$

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

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

Structure Function

Mean-reversion: a ``quantitative'' form of stationarity



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

$\kappa = \dfrac{1}{\tau}$, $\tau =$ characteristic time associated with correlation decay

AR(1) model

$$X_n = a + bX_{n-1} + \varepsilon_n \qquad \varepsilon_n \sim N\left(0, \sigma^2\right)$$

$$X_n = b^n X_0 + a\sum_{k=1}^{n} b^{n-k} + \sum_{k=1}^{n} b^{n-k}\varepsilon_k$$

$$= b^n X_0 + a\frac{b^n - 1}{b - 1} + N\left(0, \sigma^2 \frac{b^{2n} - 1}{b^2 - 1}\right)$$

$$\text{Stationarity:} \quad |b| < 1, \qquad \therefore \qquad \mu_{eq} = \frac{a}{1-b}, \quad \sigma_{eq}^2 = \frac{\sigma^2}{1-b^2}$$

$$\text{Estimation of } b: \quad \hat{b} = \frac{\sum_{t=1}^{T}\left(X_{n-t} - \bar{X}\right)\left(X_{n-t-1} - \bar{X}\right)}{\sum_{t=1}^{T}\left(X_{n-t} - \bar{X}\right)^2} \qquad \left(T = \text{time window}\right)$$

Estimation of AR(1) model

$$\varepsilon_n = X_n - a - bX_{n-1} \qquad \text{i.i.d. normals,} \quad n = 0,\ldots,T$$

$$\ln P = -\frac{1}{2\sigma^2}\sum_{n=1}^{T}\left(X_n - a - bX_{n-1}\right)^2 - \frac{T}{2}\ln\sigma^2 - \frac{T}{2}\ln(2\pi)$$

$$\left(a_{ml}, b_{ml}, \sigma_{ml}{}^2\right) = \arg\max_{a,b,\sigma^2} \ln P$$

Maximum likelihood ~
minimum least squares

$$a_{ml} = \frac{\langle X_{n+1}\rangle\langle X_n^2\rangle - \langle X_n X_{n+1}\rangle}{\langle X_n^2\rangle - \left(\langle X_n\rangle\right)^2}, \qquad b_{ml} = \frac{\langle X_n X_{n+1}\rangle - \langle X_n\rangle\langle X_{n+1}\rangle}{\langle X_n^2\rangle - \left(\langle X_n\rangle\right)^2}$$

$$\sigma_{ml}{}^2 = \left\langle \left(X_{n+1} - a_{ml} - b_{ml}X_n\right)^2 \right\rangle$$

where $\langle X_n\rangle = \frac{1}{T}\sum_{t=0}^{T-1} X_t, \quad \langle X_n\rangle = \frac{1}{T}\sum_{t=0}^{T-1} X_{t+1}$

Estimation of Ornstein-Uhlenbeck models

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

$$X_{n+1} = a + bX_n + \varepsilon_{n+1} \qquad \{\varepsilon_n\} \text{ i.i.d. } N\left(0, \sigma^2\left(\frac{1-e^{-2k\Delta t}}{2k}\right)\right)$$

$$b = \text{SLOPE}\left(\left(X_{n-l}, \ldots, X_n\right); \left(X_{n-l-1}, \ldots, X_{n-1}\right)\right),$$
$$a = \text{INTERCEPT}\left(\left(X_{n-l}, \ldots, X_n\right); \left(X_{n-l-1}, \ldots, X_{n-1}\right)\right)$$
$$k = \frac{1}{\Delta t} \ln\left(\frac{1}{b}\right), \quad m = \frac{a}{1-b}, \quad \sigma = \frac{\text{STDEV}\left(X_{n+1} - bX_n - a\right)}{\sqrt{1-b^2}} \sqrt{2 \frac{1}{\Delta t} \ln\left(\frac{1}{b}\right)}$$

Random Walks, 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

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

Fractional 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*)

Auto-regressive Models AR(m)

$X_1, X_2, \ldots, X_n, \ldots$ Time-series data to be modeled

$$X_n = a + \sum_{k=1}^{m} b_k X_{n-k} + \varepsilon_n \qquad \varepsilon_n \sim N(0, \sigma^2), \text{ i.i.d.}$$

$$Y_n^T = (X_{n-m+1}, \ldots, X_n)$$

$$B = \begin{pmatrix} b_1 & b_2 & \ldots & b_m \\ 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ & & 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} a \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \quad E_n = \begin{pmatrix} \varepsilon_n \\ 0 \\ \ldots \\ 0 \end{pmatrix}$$

$$Y_n = A + BY_{n-1} + E_n$$

$$Y_n = B^n Y_0 + \sum_{k=1}^{n} B^{n-k} A + \sum_{k=1}^{n} B^{n-k} E_k$$

AR(m) is a ``vector'' AR(1) model

Stationarity of AR(m)

$$\mu := E(X_n)$$

$$E(X_n) = a + \sum_{k=1}^{m} b_k E(X_{n-k}) \qquad \therefore \qquad \mu = a + \mu \sum_{k=1}^{m} b_k$$

$$\mu = \frac{a}{1 - \sum_{k=1}^{m} b_k} \qquad \text{necessary condition for stationarity: } \sum_{k=1}^{m} b_k \neq 1$$

$$Z_n := X_n - \mu \qquad Z_n \sim AR(m) \quad \text{with } a = 0$$

B is a contraction iff all of its eigenvalues are less than 1

$$\det(B - \lambda I) = (-1)^m \left(\lambda^m - \sum_{k=1}^{m} \lambda^{m-k} b_k \right) = (-1)^m P(\lambda)$$

All the roots of $P(\lambda)$ must satisfy $|\lambda| < 1$

ARCH(p) Errors

ARCH models the conditional variance of the innovation as a moving average of the squares of the errors in the previous terms.

$$\Phi_{n-1} = \text{ observations until time } n-1$$

$$\varepsilon_n \mid \Phi_{n-1} \sim N\left(0, \sigma_n^2\right)$$

$$\sigma_n^2 = \omega + \sum_{k=1}^{p} \alpha_k \varepsilon_{n-k}^2$$

For instance,

$$\alpha_k = \frac{1}{p}$$

Moving window of length p

$$\alpha_k = \frac{\theta^k}{\sum_{l=1}^{p} \theta^l}$$

Exponential moving average (theta<1)

GARCH(p,q) errors

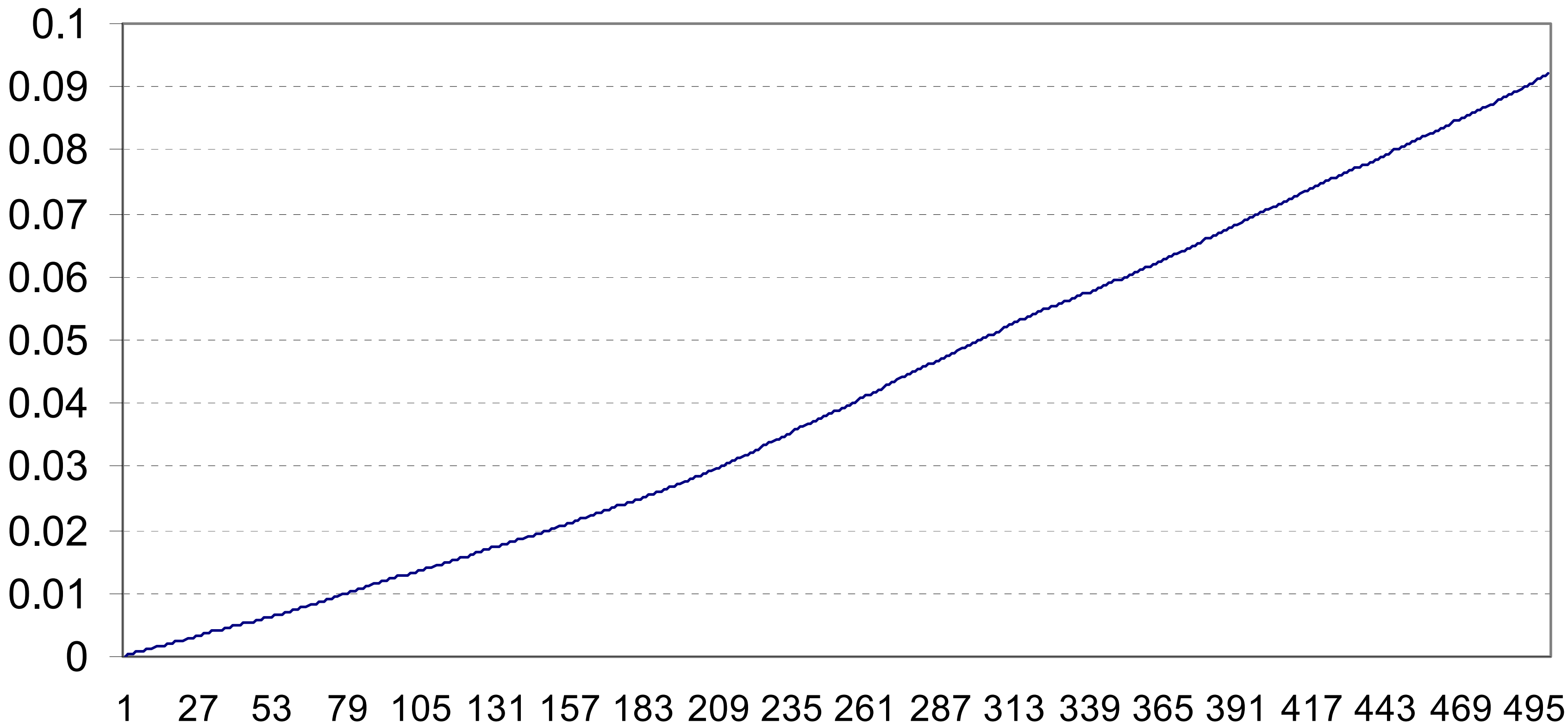
$$\varepsilon_n \mid \Phi_{n-1} \sim N\left(0, \sigma_n^2\right)$$

$$\sigma_n^2 = \omega + \sum_{k=1}^{p} \alpha_k \varepsilon_{n-k}^2 + \sum_{k=1}^{p} \beta_k \sigma_{n-k}^2$$

Generalizes ARCH by permitting the conditional volatility to be influenced by previous estimates as well as by the EMA of squared errors.

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.

Systematic Approach for looking for mean-reversion 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

More on mean-reversion model

The factors are either

A. eigenportfolios corresponding to significant eigenvalues of the market

B. industry ETF, or portfolios of ETFs (we shall use these in light of last lecture and because it's easier)

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?

Experiment: consider 39 stocks associated with XLK (SPDR Tech ETF)

Regressing returns of XYZ vs. XLK: 60-day window Betas(1/09-2/10)

	AAPL	ACS	ADBE	AKAM	APD	APH	BMC	CA	CPWR	CRM
average	1.03	0.66	1.35	1.3256	1.0962	1.28	0.72	1.04	1	1.46
stdev	0.0959	0.16	0.14	0.2081	0.141	0.137	0.08	0.16	0.1	0.31
1pct q	0.8859	0.41	1.11	0.9823	0.8528	0.989	0.53	0.79	0.68	0.86
99 pct q	1.3376	1.02	1.66	1.7451	1.3665	1.571	0.89	1.34	1.14	1.94

	CSCO	CTSH	CTXS	DELL	EBAY	EMC	ERTS	FISV	FLIR	GLW
average	1.1764	1.08	1.13	1.2785	1.1911	1.1	1.08	0.9	1.01	1.36
stdev	0.0512	0.12	0.13	0.1515	0.2107	0.065	0.11	0.11	0.13	0.14
1pct q	1.0743	0.84	0.84	1.0208	0.7202	0.944	0.88	0.7	0.67	1.13
99 pct q	1.2752	1.23	1.32	1.6897	1.5253	1.239	1.29	1.11	1.23	1.59

	GOOG	HPQ	HRB	HRS	IBM	INTU	JDSU	JNPR	MFE	MSFT
average	0.7985	1.01	0.64	0.8516	0.7245	0.711	1.68	1.39	0.89	0.93
stdev	0.1069	0.1	0.25	0.2352	0.0995	0.143	0.13	0.09	0.15	0.17
1pct q	0.644	0.86	0.2	0.4447	0.5158	0.472	1.41	1.16	0.59	0.64
99 pct q	1.0359	1.2	1.17	1.3077	0.9092	0.977	1.96	1.56	1.2	1.15

Regressing returns of XYZ vs. XLK: 60-day window Betas

	NOVL	NTAP	ORCL	QCOM	RHT	SRCL	SYMC	YHOO
average	1.0213	0.66	1.34	1.3209	1.0915	1.273	0.72	1.05
stdev	0.1471	0.17	0.17	0.2288	0.1607	0.163	0.1	0.16
1pct q	0.3472	0.4	1.02	0.9302	0.7936	0.984	0.52	0.82
99 pct q	1.3399	1.02	1.66	1.7456	1.3667	1.572	0.89	1.34

Cross-sectional statistics
for Beta variability

min stdev	0.0512	CSCO
max stdev	0.3142	CRM
min range	0.2009	CSCO
max range	1.0766	CRM

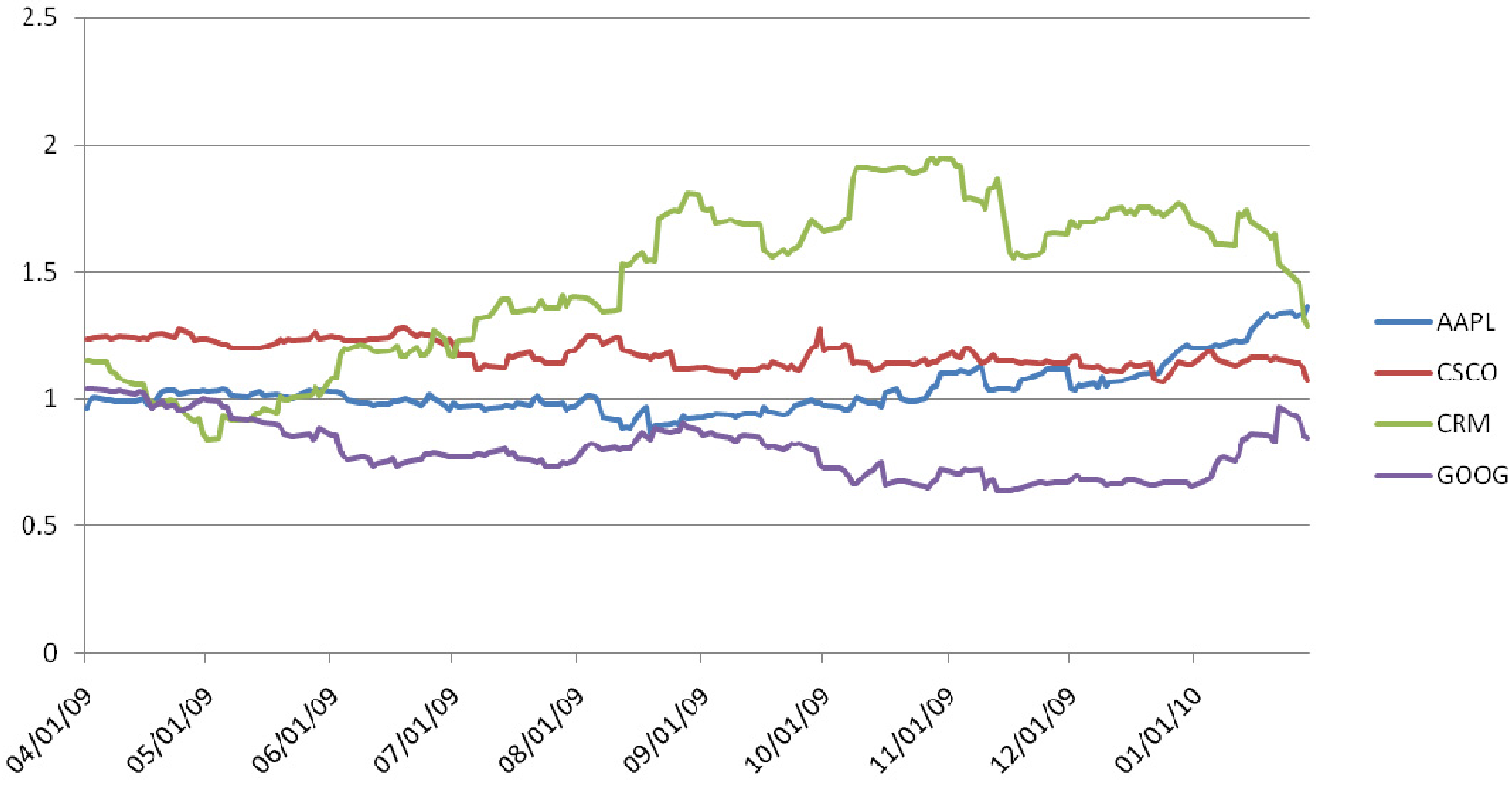
CSCO vs. XLK


Feb 23, 2010 : CSCO 24.05 XLK 21.51
60%
50%
40%
30%
20%
10%
0%
© 2008 Yahoo! Inc.
2009
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2010
Feb
Volume 36,286,000
100M
50M
1D 5D 1M YTD 3M 6M 1Y 2Y 5Y Max
FROM: Feb 24 2009 TO: Feb 23 2010
1990
1995
2000
2005
2010

CRM vs. XLK


Feb 18, 2010 : CRM 68.45 XLK 21.88
160%
140%
120%
100%
80%
60%
40%
20%
0%
2009
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2010
Feb
Volume 2,090,400
8M
6M
4M
2M
1D 5D 1M YTD 3M 6M 1Y 2Y 5Y Max
FROM: Feb 24 2009 TO: Feb 23 2010
2004
2005
2006
2007
2008
2009
2010

Evolution of 60-day Betas versus XLK: AAPL, CSCO, CRM,GOOG



Computing the residuals in practice

$X_1,....,X_T$ etf returns

$Y_1,.....,Y_T$ stock returns

$w =$ estimation window (in days)

$$\beta_{t-w,t} = SLOPE((X_{t-w},....,X_{t-1}),(Y_{t-w},....,Y_{t-1}))$$

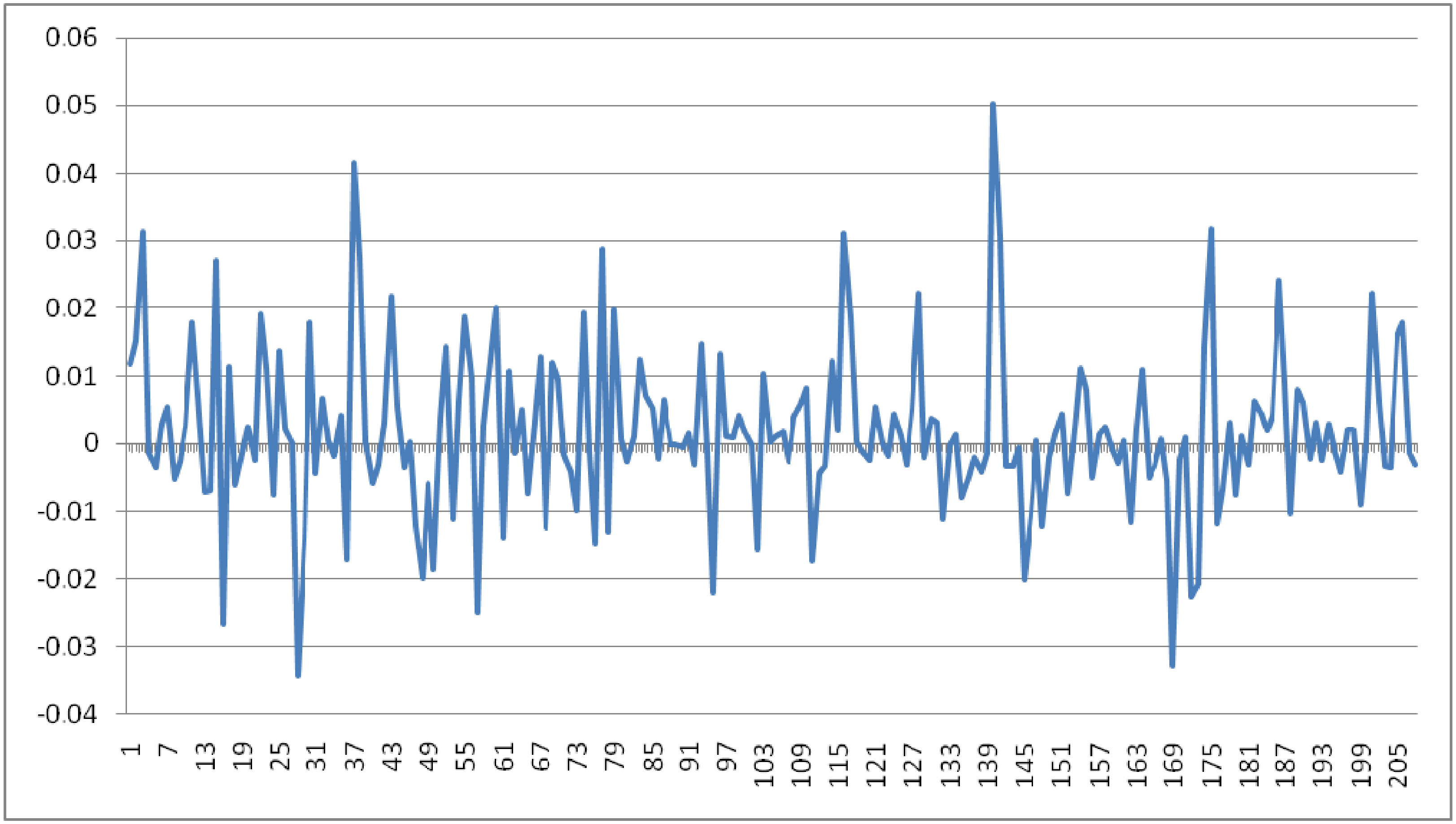
$$\varepsilon_t = Y_t - \beta_{t-w,t} X_{t,} \qquad t = w+1, w+2,...,T$$

Use window of w days before current date

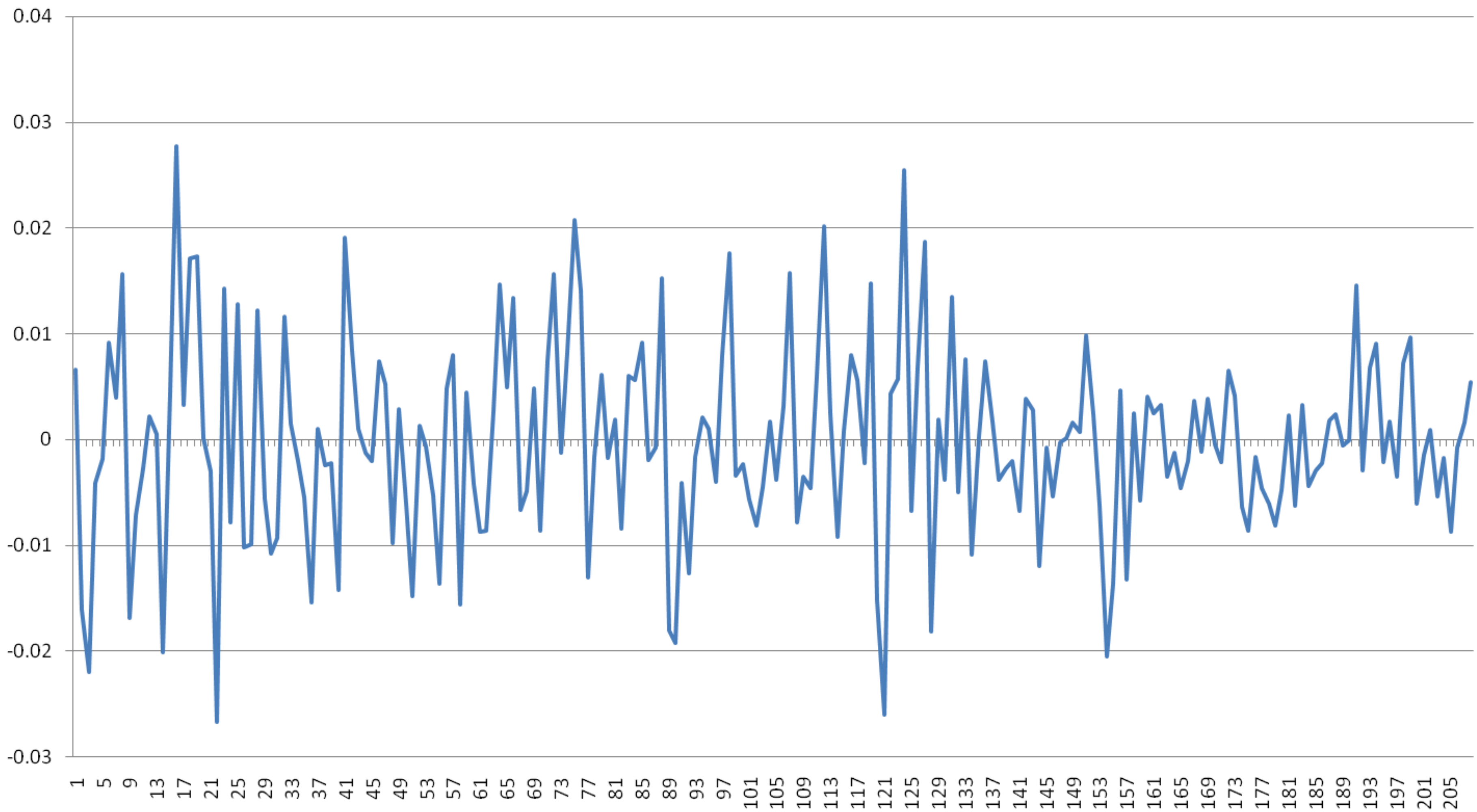
Define $$Z_t := \sum_{k=w+1}^{t} \varepsilon_k$$

``Co-integrated'' residual (CR)

AAPL Residuals (against 60-day Betas)



CSCO residuals against 60-day Betas



Computing Mean-reversion from 10/30/09 to 1/28/10

Slope b is computed using lagged regression of CR

Ticker	AAPL	ACS	ADBE	AKAM	APD	APH	BMC	CA	CPWR	CRM
b (slope)	0.95	0.98	0.88	0.87	0.94	0.86	0.76	0.89	0.97	0.85
kappa	11.9	4.43	30.9	35.3	15.7	36.9	67.7	28.3	8.89	41.7
tao(in days)	21.1	56.8	8.15	7.13	16	6.83	3.72	8.91	28.3	6.04

Ticker	CSCO	CTSH	CTXS	DELL	EBAY	EMC	ERTS	FISV	FLIR	GLW
b (slope)	0.93	0.76	1.02	0.93	0.8	0.82	0.89	0.9	0.92	0.97
kappa	17.3	68	-5.7	17.4	56.1	49.4	28.1	27.3	21.6	7.4
tao(in days)	14.6	3.71	-44.2	14.5	4.49	5.1	8.98	9.22	11.7	34.1

Ticker	GOOG	HPQ	HRB	HRS	IBM	INTU	JDSU	JNPR	MFE	MSFT
b (slope)	0.96	0.81	0.97	0.88	0.7	0.87	0.91	0.88	0.93	0.75
kappa	10.9	52.5	8.66	32.5	91.3	36.4	25.1	31.3	17.8	73.2
tao(in days)	23.1	4.8	29.1	7.75	2.76	6.93	10	8.04	14.2	3.44

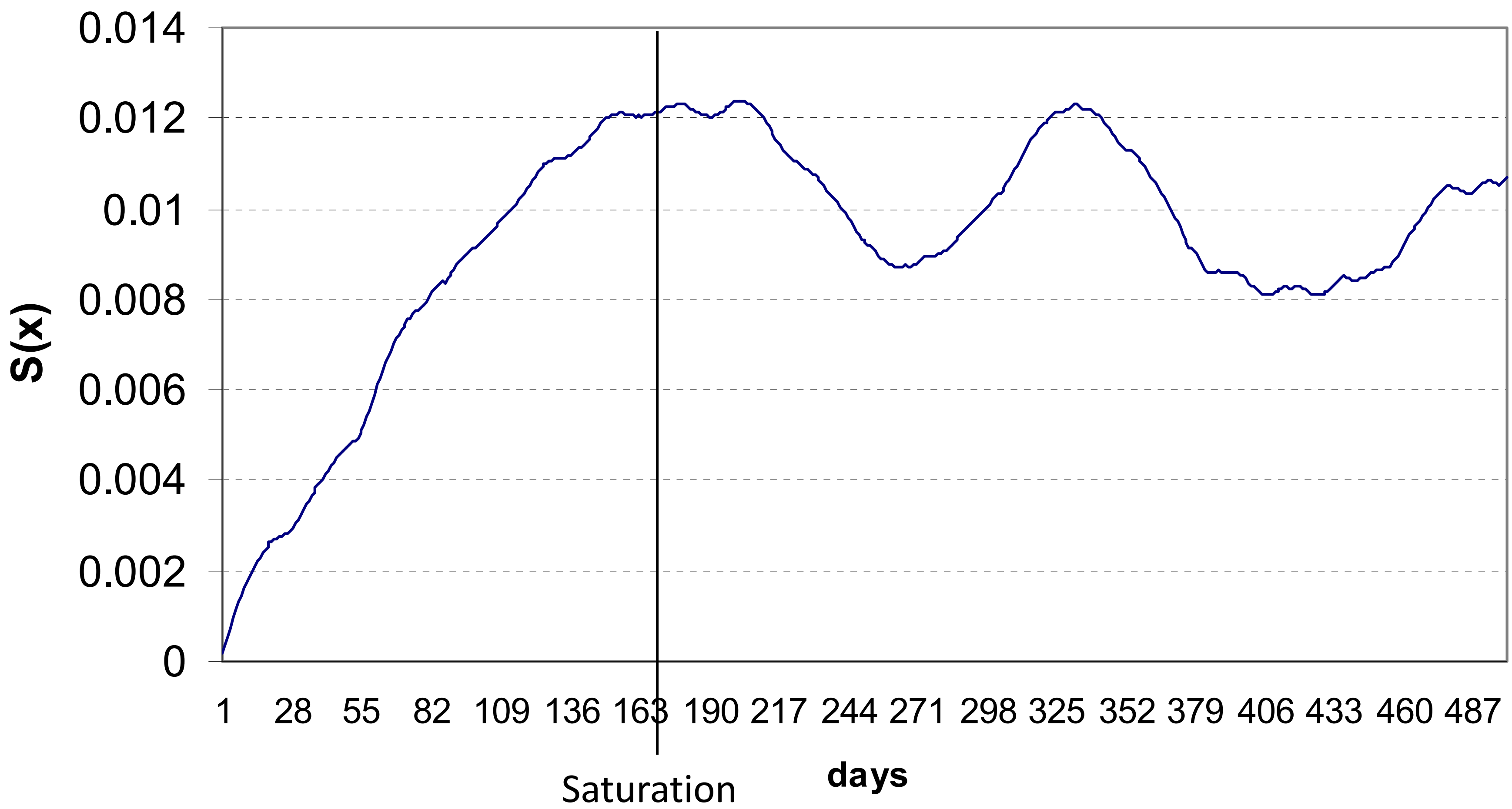
Computing Mean-reversion from 10/30/09 to 1/28/10

Ticker	NOVL	NTAP	ORCL	QCOM	RHT	SRCL	SYMC	YHOO
b (slope)	0.98	0.9	0.96	0.78	0.88	0.91	0.88	0.91
kappa	5.48	25.8	9.28	61.7	33.4	22.9	31.2	25
tao(in days)	46	9.78	27.2	4.08	7.55	11	8.08	10.1

Stocks with mean-reversion of less than 10 days

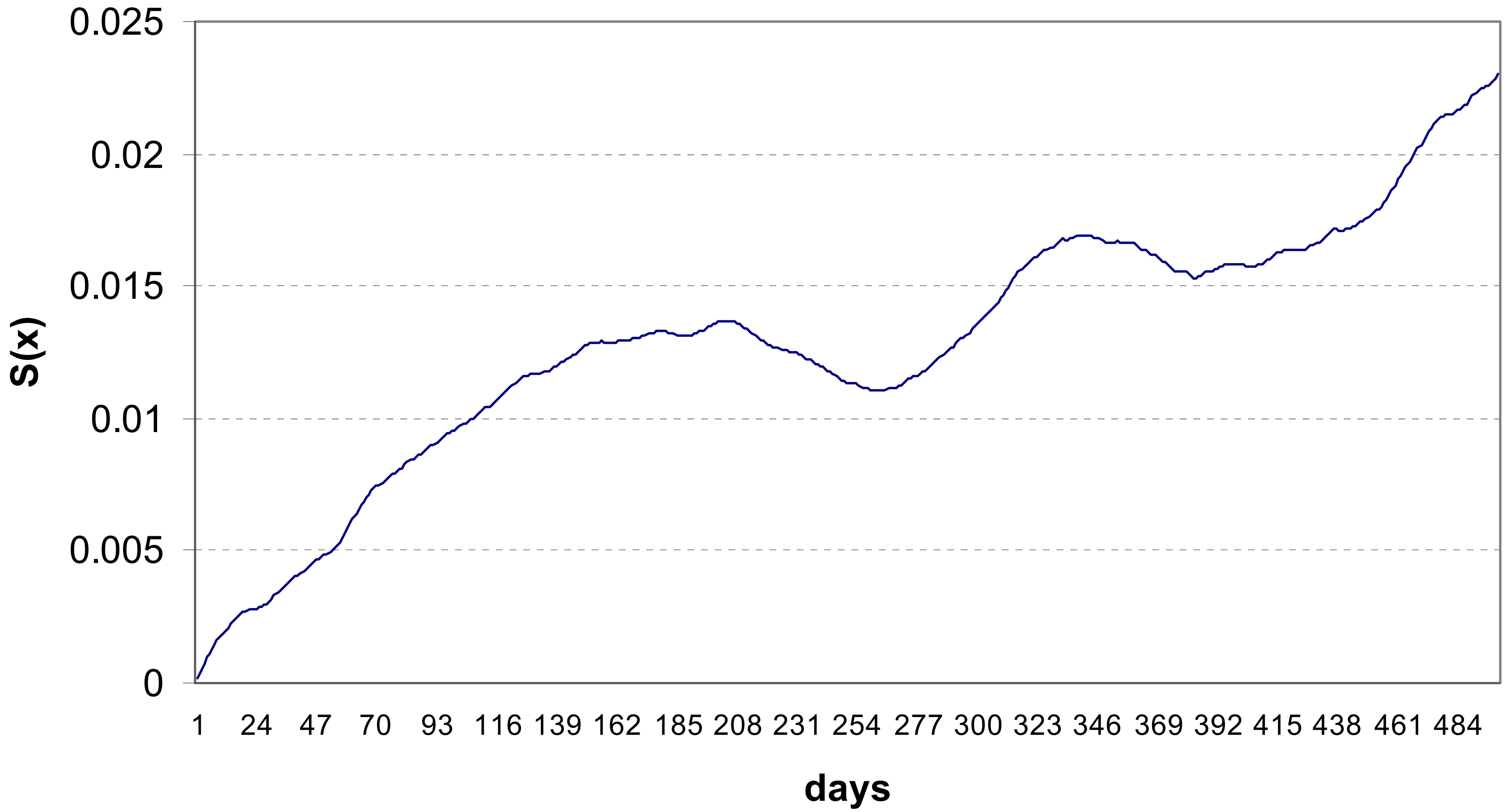
stock	b(slope)	kappa	tao(days)
IBM	0.70	91	2.8
MSFT	0.75	73	3.4
CTSH	0.76	68	3.7
BMC	0.76	68	3.7
QCOM	0.78	62	4.1
EBAY	0.80	56	4.5
HPQ	0.81	53	4.8
EMC	0.82	49	5.1
CRM	0.85	42	6
APH	0.86	37	6.8
INTU	0.87	36	6.9
AKAM	0.87	35	7.1
RHT	0.88	33	7.6
HRS	0.88	33	7.8
JNPR	0.88	31	8
SYMC	0.88	31	8.1
ADBE	0.88	31	8.1
CA	0.89	28	8.9
ERTS	0.89	28	9
FISV	0.90	27	9.2
NTAP	0.90	26	9.8

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



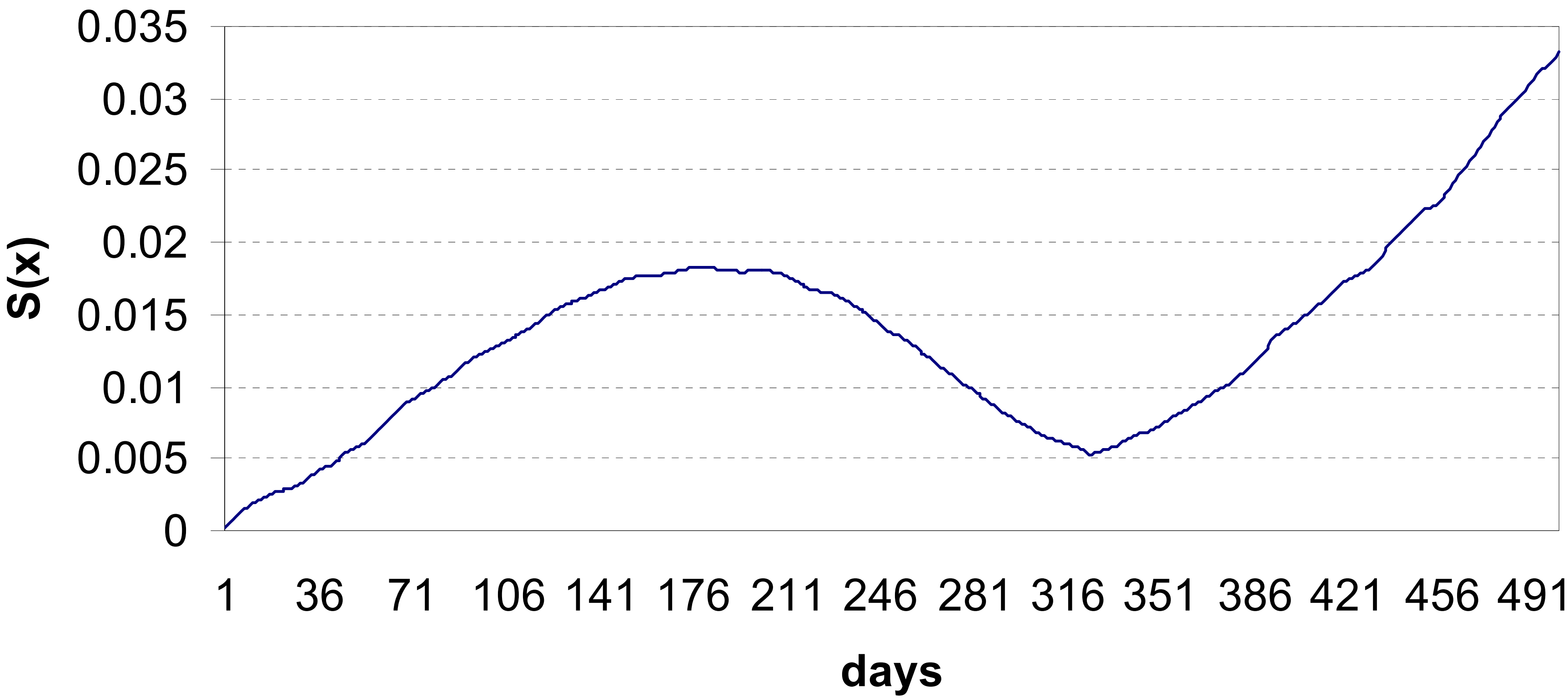
OIH: Oil Services ETF, SLB: Schlumberger-Doll Research

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



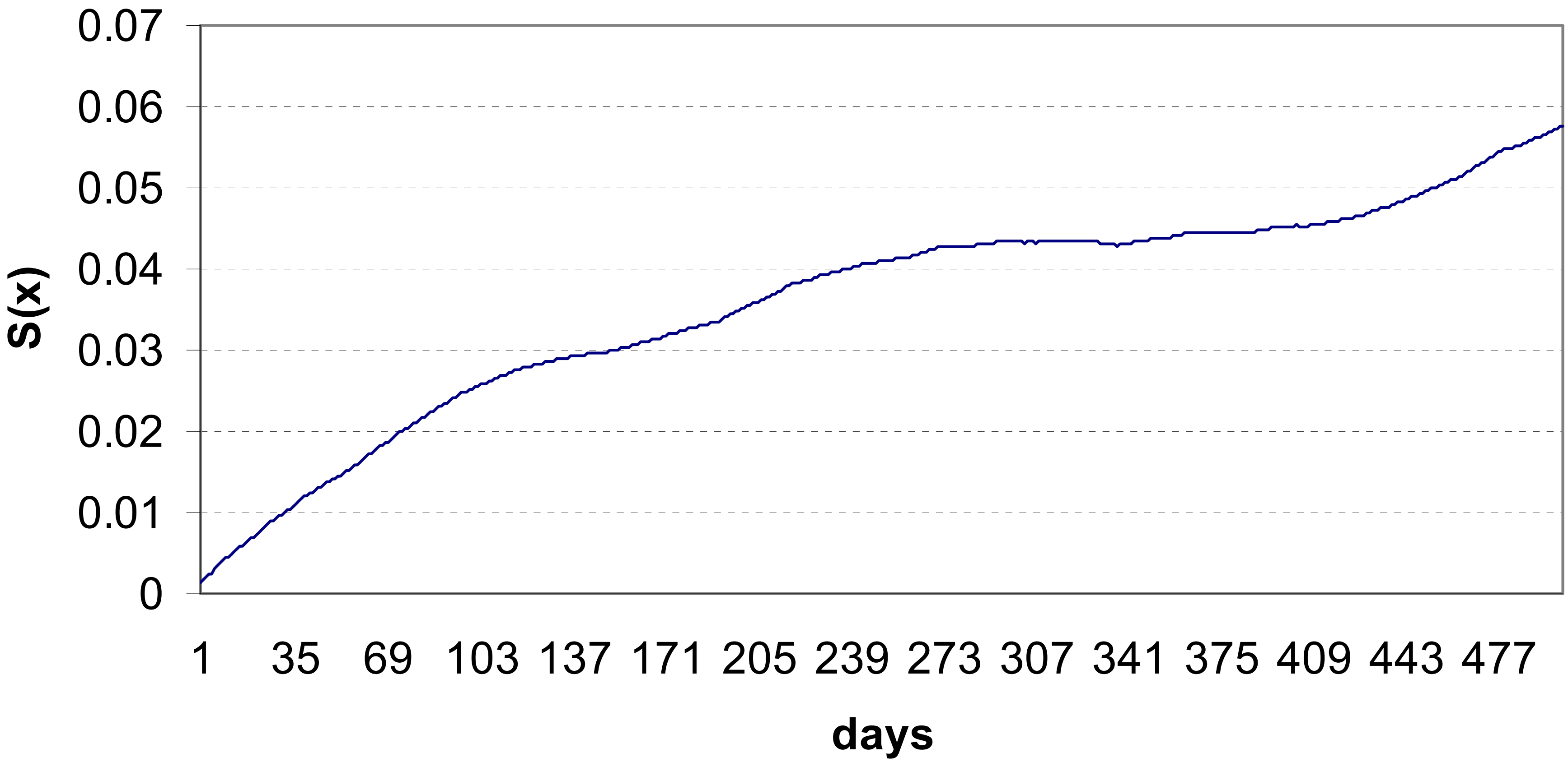
$$P_{n+1} = P_n \times \left(1 + R_{\text{slb}} - R_{\text{oih}}\right), \quad X_n = \ln P_n$$

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

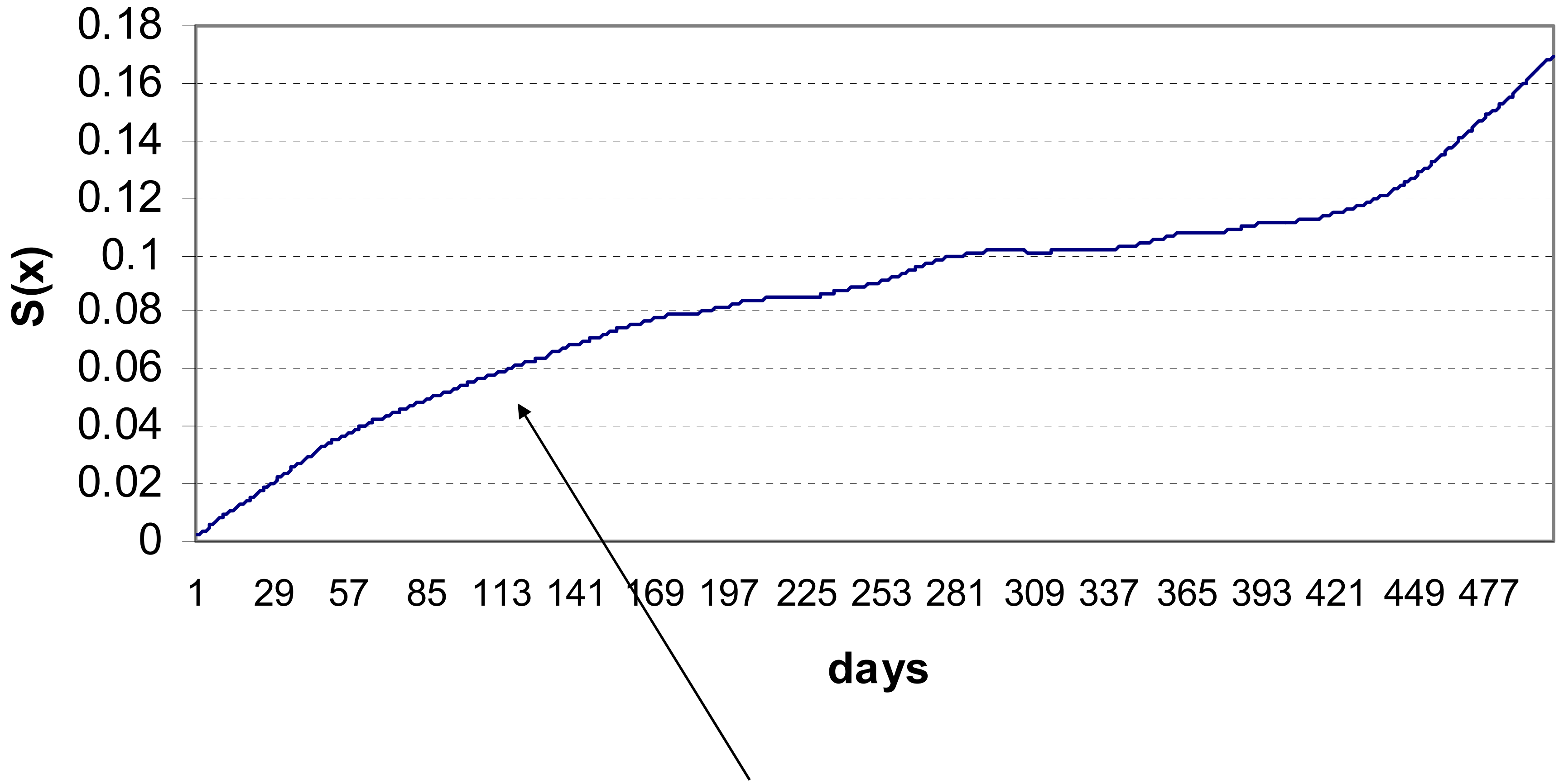


$$P_{n+1} = P_n \times \left(1 + R_{\mathrm{slb}} - \beta_{60\mathrm{d}} \cdot R_{\mathrm{oih}}\right), \quad X_n = \ln P_n$$

Structure Function log (GENZ/IBB)

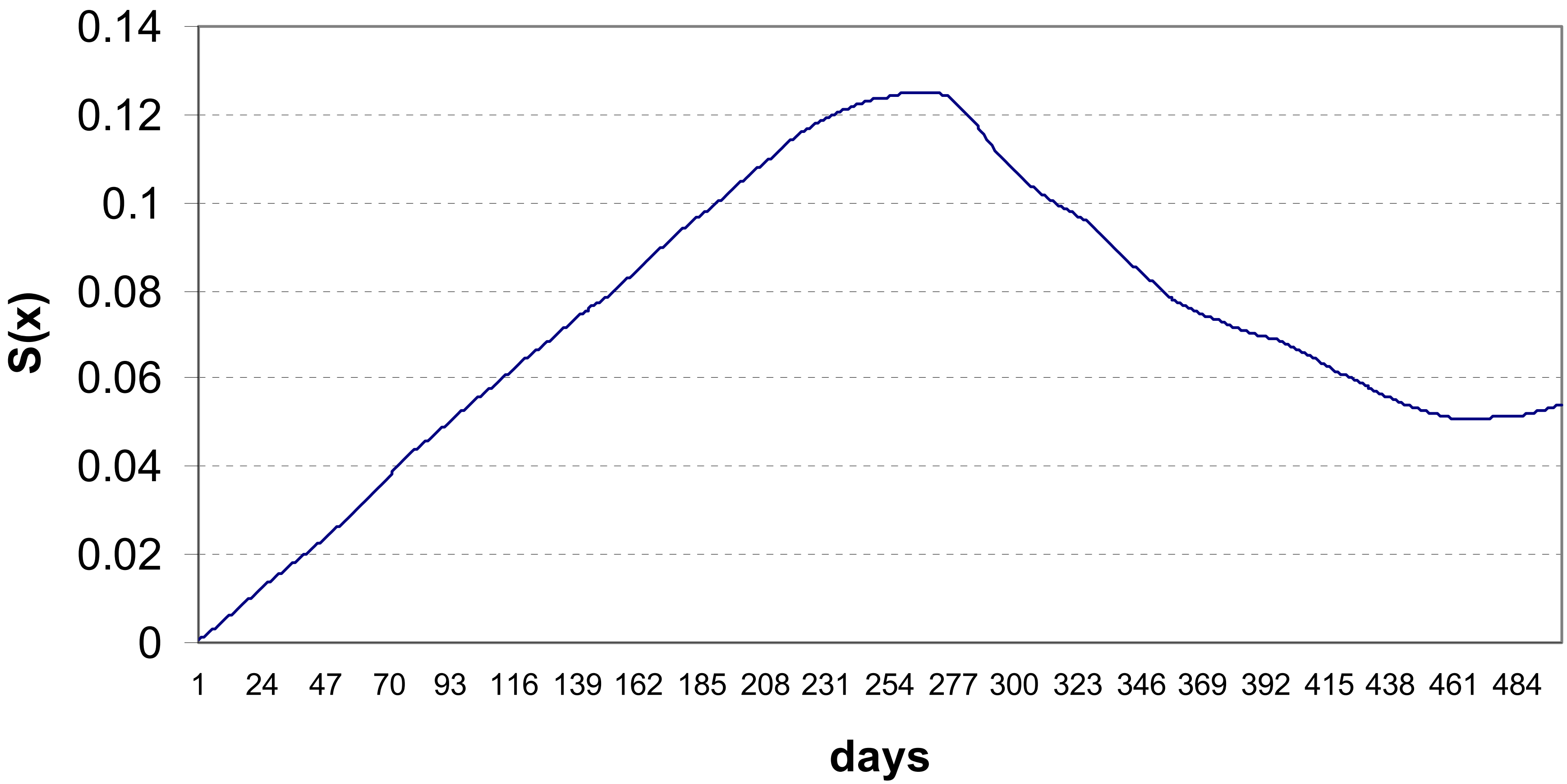


Structure function ln (DNA/GENZ)



DNA: Genentech Inc.
GENZ; Genzyme Corp.

Structure Fn for Beta-Neutral GENZ-DNA Spread



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