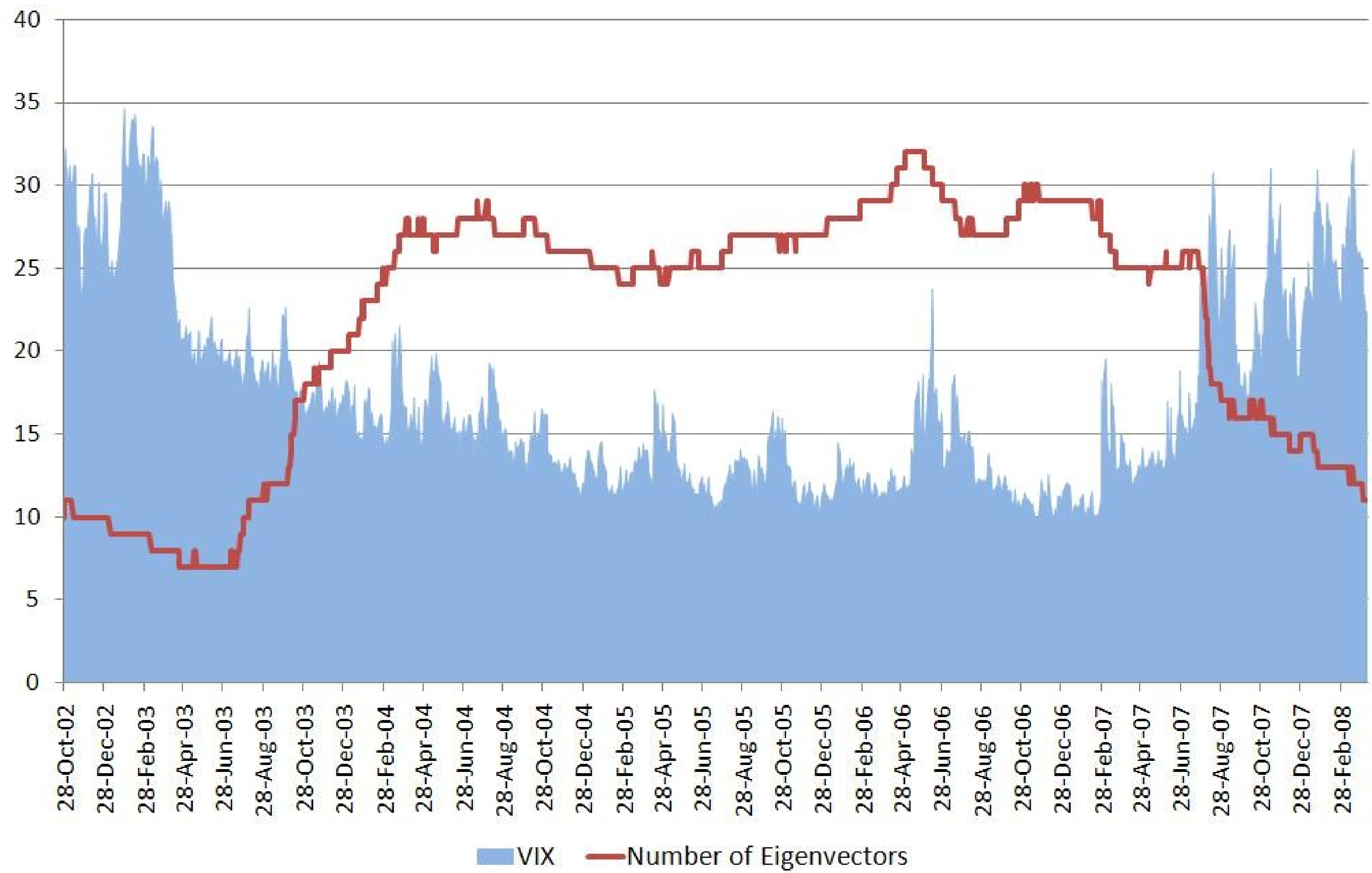
Risk and Portfolio Management
Spring 2011

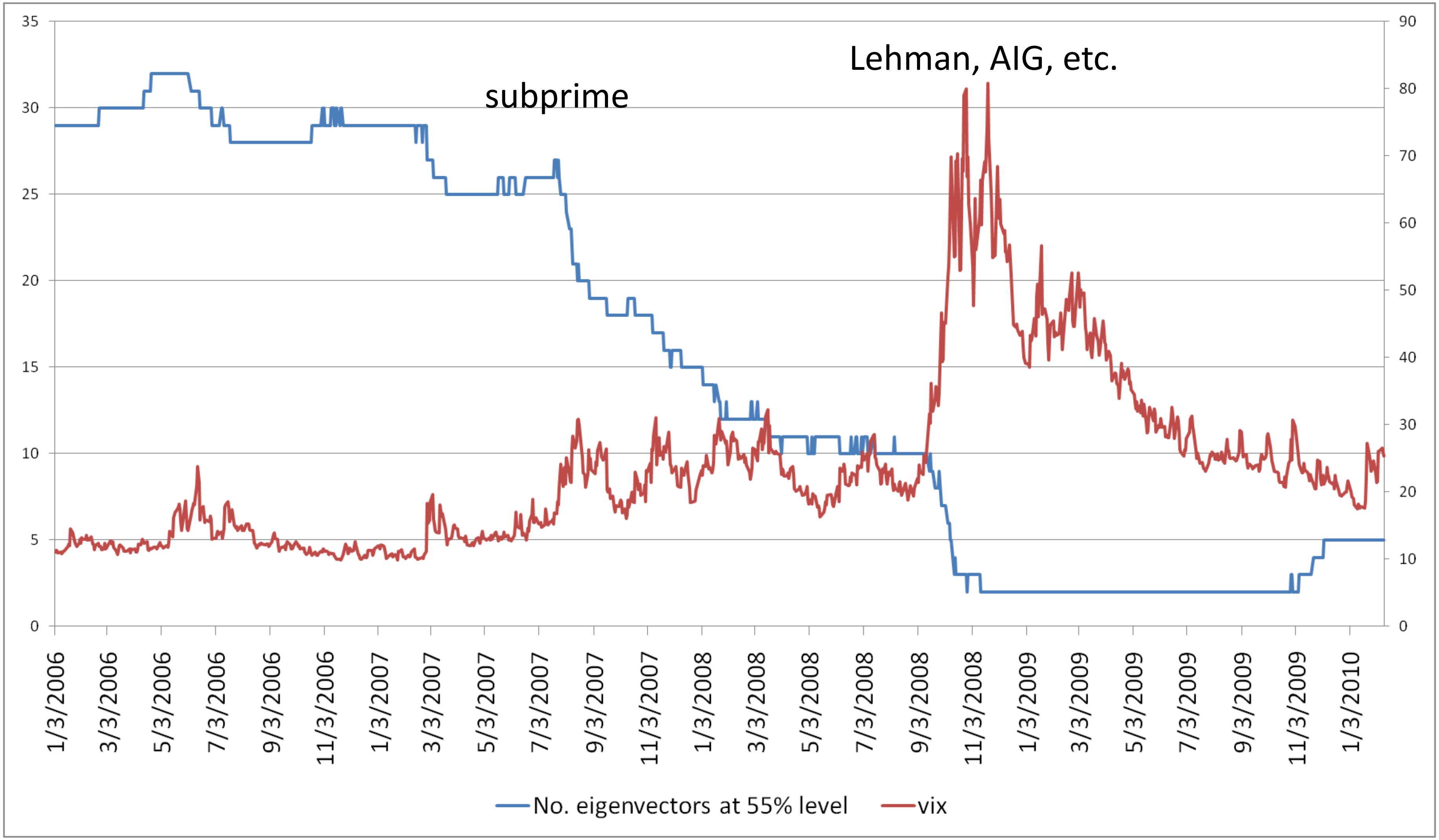
Econometric models & stochastic processes

for prices, volatilities, spreads…

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



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

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

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

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

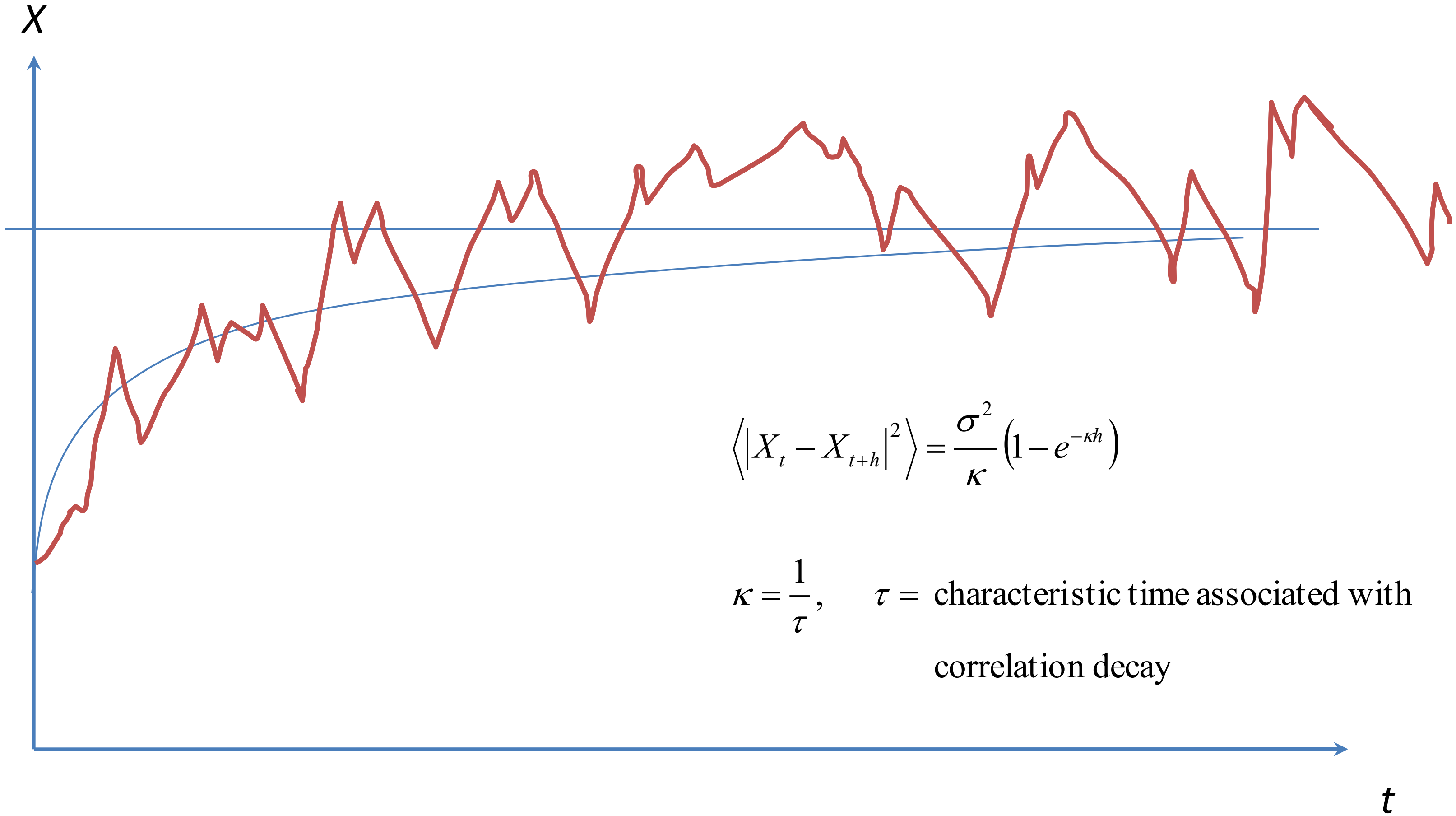
$$= \sigma^2 e^{-kh} \int_{-\infty}^{t} e^{-2k(t-s)} ds$$

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

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

Structure Function

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



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

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

AR(1) model

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

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

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

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

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

Estimation of AR(1) model

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

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

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

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

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

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

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

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

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

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)

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

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

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

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

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 $|\lambda| < 1$

Auto-regressive Models

ARCH(p), GARCH(p,q)

Following R. Engle and T. Bollerslev

Conditional Mean and Conditional Variance

$$y_t, \quad t = 1,2,3,\ldots,T$$

Given time series

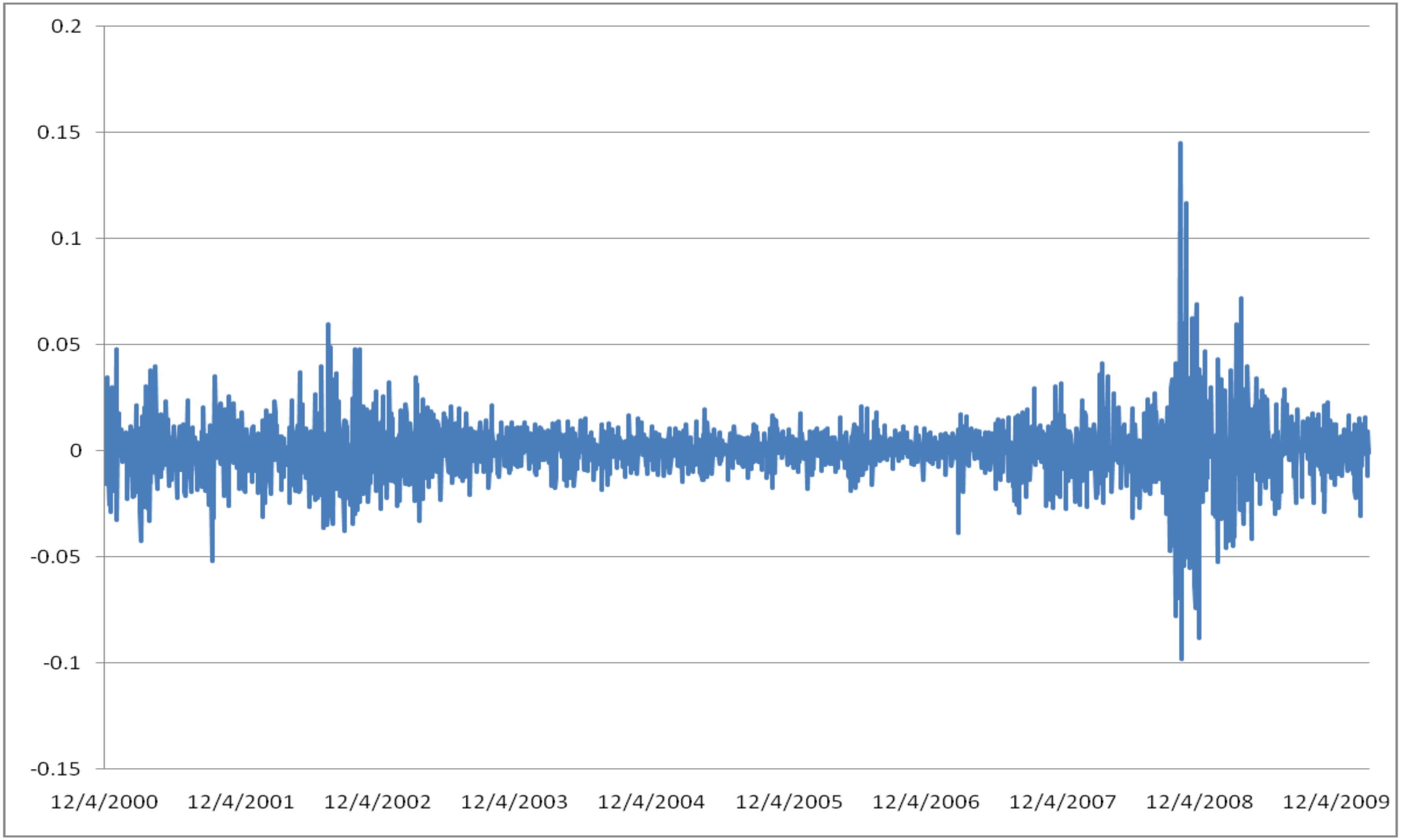
$$p(y_t \mid y_{t-1}, y_{t-2}, \ldots) = p(y_t \mid \Phi_{t-1})$$

Model the <u>conditional distributions</u>

$$y_t = \mu(\Phi_{t-1}) + \sigma(\Phi_{t-1})\varepsilon_t, \quad E(\varepsilon_t) = 0, \; E(\varepsilon_t^2) = 1$$

$$\text{Example:} \quad y_t \mid \Phi_{t-1} \sim N\left(\mu(\Phi_{t-1}), \sigma^2(\Phi_{t-1})\right)$$

Returns of S&P 500 Index 12/1/2000-2/26/2010



ARCH(p) (Engle, 1982)

$$y_t = \alpha + \beta x_t + u_t$$

Uncorrelated residuals does not necessarily imply independent residuals

$$u_t = h_t^{1/2} \varepsilon_t \qquad E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1$$

$$h_t = a_0 + a_1 u_{t-1}^2$$

Unlike in AR, the error is not assumed to have constant variance.

More generally,

$$h_t = a_0 + \sum_{k=1}^{p} a_k u_{t-k}^2$$

Conditional variance is a lagged sum of squared residuals, eg.

$$h_t = \frac{1}{T} \sum_{k=1}^{T} u_{t-k}^2$$

GARCH(p,q) (Bollerslev, 1986)

$$u_t = h_t^{1/2}\varepsilon_t \qquad E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1$$

$$h_t = \omega + \sum_{i=1}^{p}\alpha_i u_i^2 + \sum_{j=1}^{q}\beta_j h_{t-j}$$

Dependence on previous squared returns and previous conditional variances.

Most famous versions in practice: GARCH(1,1) or GARCH (1,p) which are basically AR(p) processes on the conditional variance driven by the squared-returns process

GARCH(1,1)

$$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}$$

1-lag dependence

$$\begin{aligned} h_t &= \omega + \alpha u_{t-1}^2 + \beta\left(\omega + \alpha u_{t-2}^2 + \beta h_{t-2}\right) \\ &= \omega + \beta\omega + \alpha\left(\beta u_{t-2}^2 + u_{t-1}^2\right) + \beta^2 h_{t-2} \\ &\ddots \end{aligned}$$

$$\boxed{h_t = \frac{\omega}{1-\beta} + \alpha \sum_{k=1}^{\infty} \beta^k u_{t-k}^2}$$

GARCH(1,1) is an exponentially weighted moving average of squared-errors. Beta determines the effective ``window size'' for estimation of conditional variance.

GARCH(1,2)

$$\begin{pmatrix} h_t \\ h_{t-1} \end{pmatrix} = \begin{pmatrix} \omega \\ 0 \end{pmatrix} + \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_{t-1}^2 \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h_{t-1} \\ h_{t-2} \end{pmatrix}$$

Vector AR(1)

Stability condition: $\lambda^2 - \beta_1 \lambda - \beta_2 = 0 \Rightarrow |\lambda| < 1$

$$h_t = \bar{h} + A \sum_{k=1}^{\infty} \lambda_1^k u_{t-k}^2 + B \sum_{k=1}^{\infty} \lambda_2^k u_{t-k}^2$$

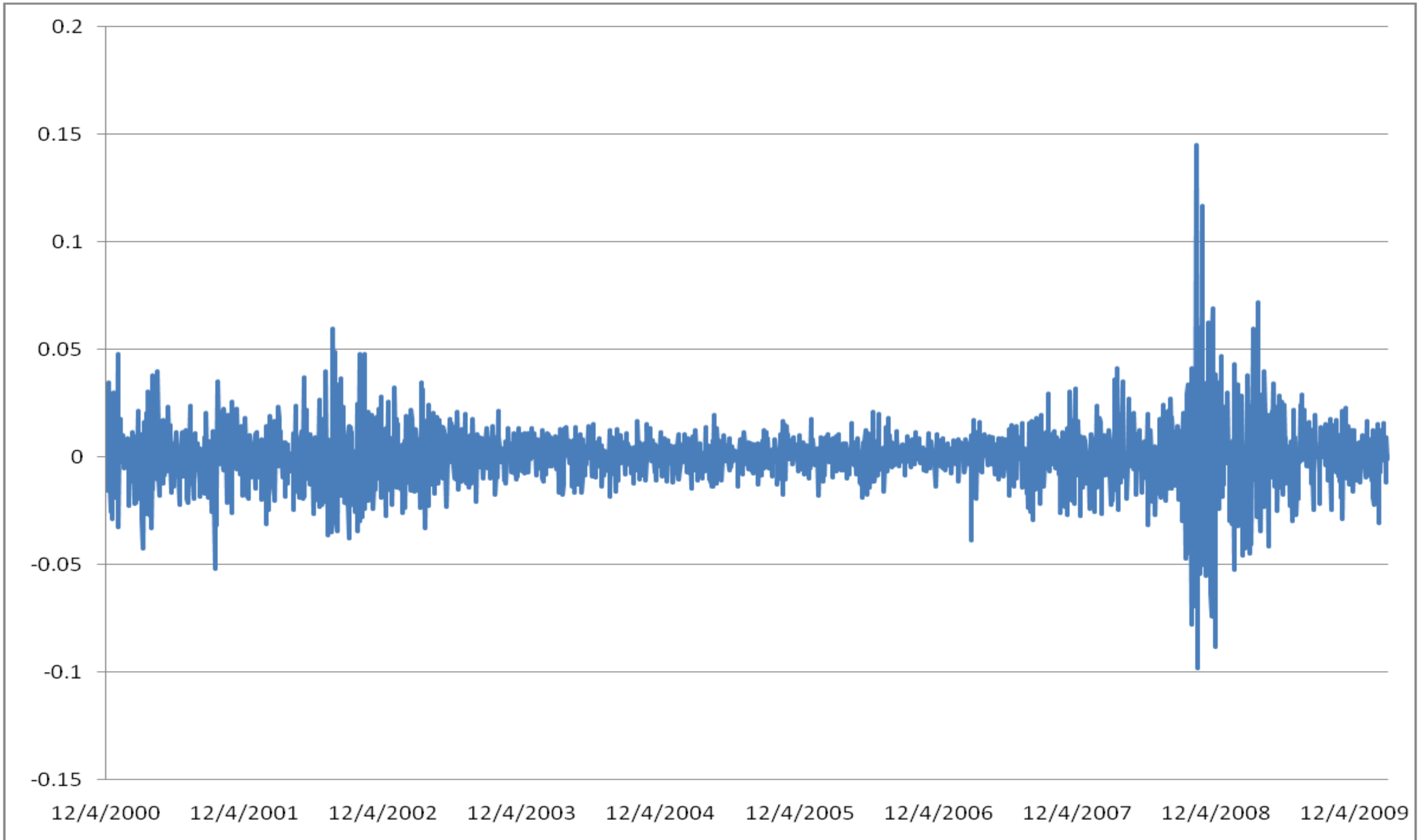
Steady-state solution

Intuitively, GARCH(1,2) is the sum of two EWMA with different time-scales (decay rates).

Notice however that the right-hand side depends on *h* as well, so the PDF of the conditional variance is not a chi-squared.

GARCH(1,p) is the sum of (at most) p EWMAs.

Returns of S&P 500 Index 12/1/2000-2/26/2010



Fitting to GARCH(1,p)

We know that the tails of SPY are heavy and behave like Student t with df~3.5

This heavy-tailed behavior of stock prices can be modeled by assuming a static distribution (Student) or a time-dependent distribution with a GARCH-type stochastic conditional variance.

The latter approach (GARCH) has the advantage that it incorporates dynamics so it may capture ``persistence'' of volatility across time.

From a portfolio risk-management perspective, the situation is ``cured'' by assuming a Student-t distribution with 3.5 degrees of freedom for returns (to capture tail behavior) and an EWMA variance which is adjusted daily to capture volatility clustering effects.

The question that remains is: what is the correct estimation window?

GARCH(1,1) estimation of SPY returns

Method: ML - BFGS with analytical gradient

date: 03-02-10

time: 18:10

Included observations: 2320

Convergence achieved after 56 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
omega	2.85989E-06	3.9342E-07	7.269290633	3.61489E-13
alpha_1	0.698241421	0.020073908	34.78353205	0
beta_1	0.508888808	0.050794297	10.01862092	0
Log Likelihood	7053.473574			
Jarque Bera	12844.90612	Prob		0
Ljung-Box	65535	Prob		65535

GARCH(2,1) estimation

Method: ML - BFGS with analytical gradient

date: 03-03-10

time: 13:25

Included observations: 2320

Convergence achieved after 45 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
omega	2.69557E-05	2.4E-06	11.25236	0
alpha_1	0.541398855	0.073788	7.337198	2.1805E-13
alpha_2	0.355438292	0.035892	9.90302	0
beta_1	0.268210539	0.045356	5.913404	3.3511E-09

Log Likelihood	7060.668319			
Jarque Bera	12844.90612	Prob		0
Ljung-Box	65535	Prob		65535

Garch(1,2)

Method: ML - BFGS with analytical gradient

date: 03-03-10

time: 13:34

Included observations: 2320

Convergence achieved after 54 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
omega	1.93253E-06	3.45079E-07	5.600257981	2.14033E-08
alpha_1	0.347594236	0.053959618	6.441747563	1.18106E-10
beta_1	0.417978993	0.040988575	10.19745117	0
beta_2	0.329591408	0.064169394	5.136271201	2.80243E-07

Log Likelihood	7119.174476			
Jarque Bera	12844.90612	Prob		0
Ljung-Box	65535	Prob		65535

Which model should we use?

All three GARCH models fit the data very well, with high z-statistics.

Preference should be given to the model with smallest number of parameters, so GARCH(1,1) should be suitable.

Cointegration and Pairs Trading

In the previous lecture we saw some examples of pairs trading with ETFs

X_t = return on XLK

Y_t = return on EBAY

Perform $m-$day regression to construct residuals

$$Y_t = \beta X_t + \varepsilon_t$$

$$\beta = \mathrm{SLOPE}\left(\left(Y_{t-m},\ldots,Y_{t-1}\right),\left(X_{t-m},\ldots,X_{t-1}\right)\right)$$

$$\varepsilon_t = Y_t - \beta X_t$$

$$\mathrm{P\,\&\,L} = 100 * \prod_{k=1}^{t}\left(1+\varepsilon_k\right) \qquad y_t = y_0 + \sum_{k=1}^{t}\ln\left(1+\varepsilon_k\right)$$

Question of interest : is y_t stationary? Does y_t have a `unit root'?

Dickey-Fuller Test for Unit Roots (aka Augmented Dickey-Fuller test)

The Dickey-Fuller test is used to test for unit roots in statistical data.

Consider the following model for the differentiated time-series:

$$\Delta y_t = \alpha + \beta t + \delta_0 y_{t-1} + \sum_{k=1}^{n} \delta_k \Delta y_{t-k} + \varepsilon_t, \quad \Delta y_t = y_t - y_{t-1}$$

Null hypothesis: there is a unit root, i.e. $\delta_0 = 0.$ $\quad DF = \dfrac{\hat{\delta}_0}{\mathrm{stdev}\left(\hat{\delta}_0\right)}$

n is determined dynamically as part of the test (Akaike Information Criterion)

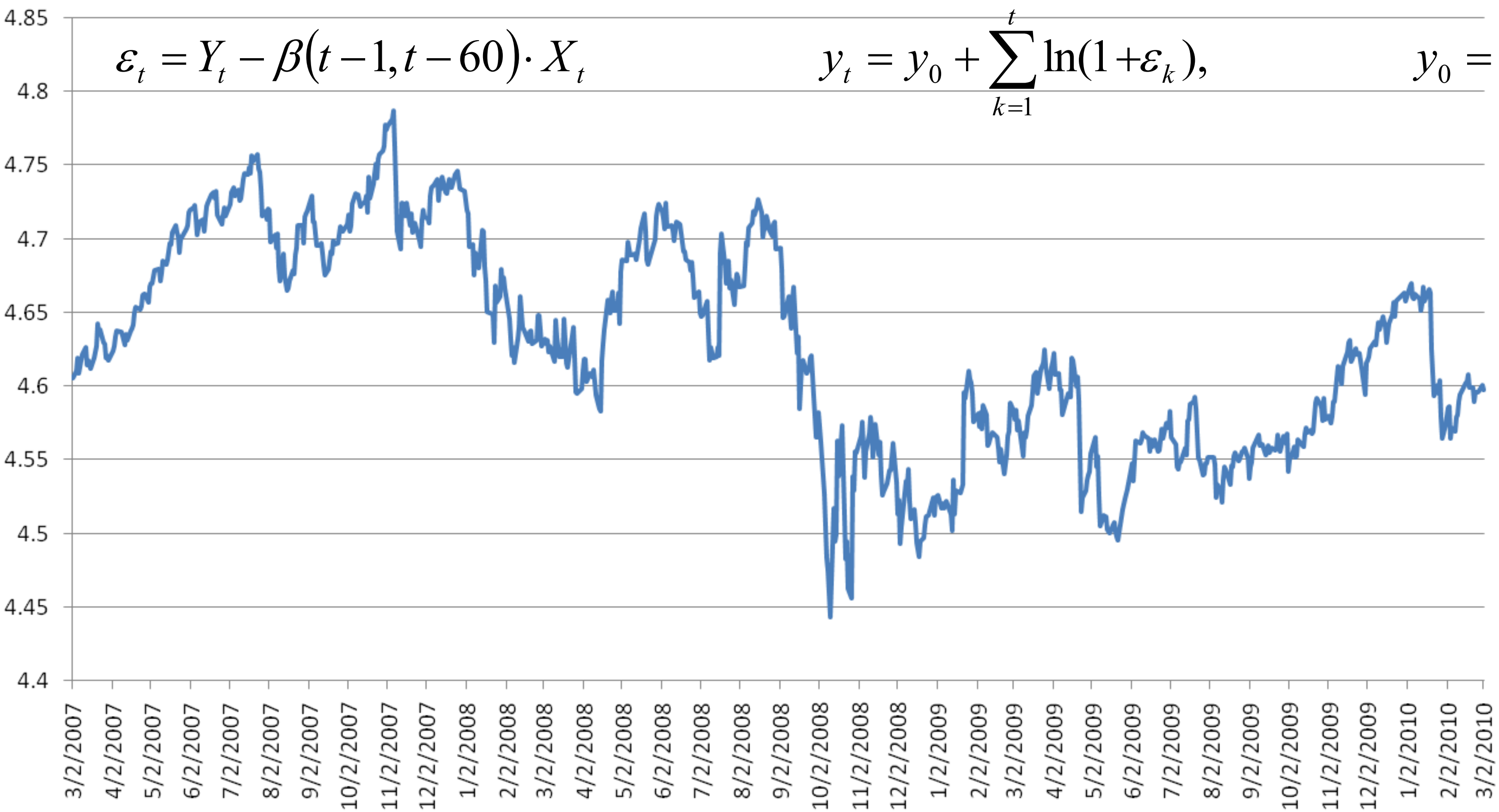
ADF Critical Values: Reject delta=0 if DF <	
1% level	-3.970385
5% level	-3.415895
10% level	-3.130187

EBAY vs. XLK residuals

$Y_t =$ daily return of EBAY

$X_t =$ daily return of XLK

$$\varepsilon_t = Y_t - \beta(t-1, t-60) \cdot X_t \qquad y_t = y_0 + \sum_{k=1}^{t} \ln(1+\varepsilon_k), \qquad y_0 = 100$$



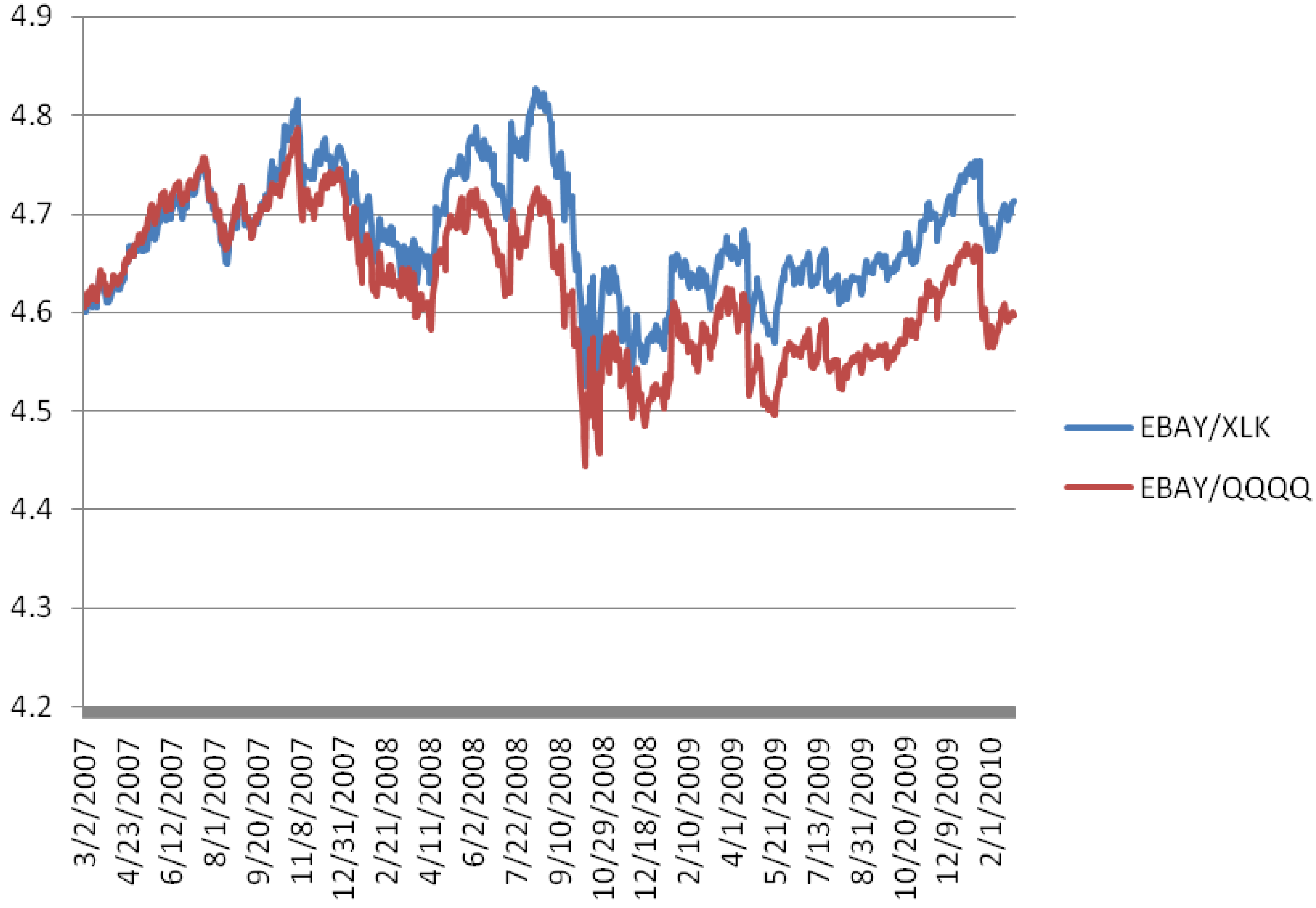
Augmented DF test for EBAY/XLK

Best lag fit: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob
tseries(-1)	-0.025582	0.009132	-2.801401	0.005222
D(tseries(-1))	-0.104975	0.036984	-2.838362	0.004660
D(tseries(-2))	0.032844	0.037145	0.884206	0.376875
D(tseries(-3))	0.041696	0.036765	1.134124	0.257113
D(tseries(-4))	-0.139433	0.036498	-3.820306	0.000145
D(tseries(-5))	0.023322	0.036852	0.632844	0.527033
D(tseries(-6))	-0.103297	0.036384	-2.839106	0.004649
D(tseries(-7))	-0.123580	0.036566	-3.379630	0.000764
D(tseries(-8))	0.062589	0.036842	1.698850	0.089771
D(tseries(-9))	0.103669	0.036604	2.832135	0.004751
C	0.120657	0.043010	2.805296	0.005160
@trend	-0.000006	0.000003	-2.076142	0.038228

Cannot reject UR @ 90% level

EBAY vs. QQQQ residuals



ADF for EBAY/QQQQ

Null Hypothesis: tseries has a unit root

Exogenous: Constant and linear Trend

Lag Length: 4 (Automatic Based on AIC, MAXLAG=10)

Variable	Coefficient	Std. Error	t-Statistic	Prob
tseries(-1)	-0.023280	0.008338	-2.791940	0.005374
D(tseries(-1))	-0.078624	0.036419	-2.158873	0.031179
D(tseries(-2))	0.019488	0.036533	0.533428	0.593897
D(tseries(-3))	0.030306	0.036525	0.829726	0.406960
D(tseries(-4))	-0.114959	0.036359	-3.161785	0.001632
C	0.109870	0.039251	2.799187	0.005256
@trend	-0.000002	0.000002	-0.918302	0.358759

ARMA(p,q) process

$$y_t = a_0 + \sum_{k=1}^{p} a_k y_{t-k} + \sum_{l=1}^{q} b_k u_{t-k} + u_t$$

Combines autorregressive models with moving average models

Simple linear time-series model

Fitting to an ARMA(1,1)

timeseries: y
Method: Nonlinear Least Squares (Levenberg-Marquardt)
date: 03-03-10 time: 18:52
Included observations: 755
p = 1 - q = 1 - constant - manual selection

	Coefficient	Std. Error	t-Statistic	Prob.
c	4.627335411	0	148.9024	0
AR(1)	0.986154258	0	159.9401	0
MA(1)	-0.110605985	0	-2.998961	0.002798377

R-squared	0.965239	Mean dependent var	4.628068
Adjusted R-squared	0.965147	S.D. dependent var	0.071188
S.E. of regression	0.013290	Akaike info criterion	-5.791955
Sum squared resid	0.132821	Schwarz criterion	-5.773571
Log likelihood	2189.462984	Durbin-Watson stat	2.007356

Inverted AR-roots	0.99
Inverted MA-roots	0.11

Fitting y to an AR(1) process

timeseries: y

Method: Nonlinear Least Squares (Levenberg-Marquardt)

date: 03-03-10 time: 18:49

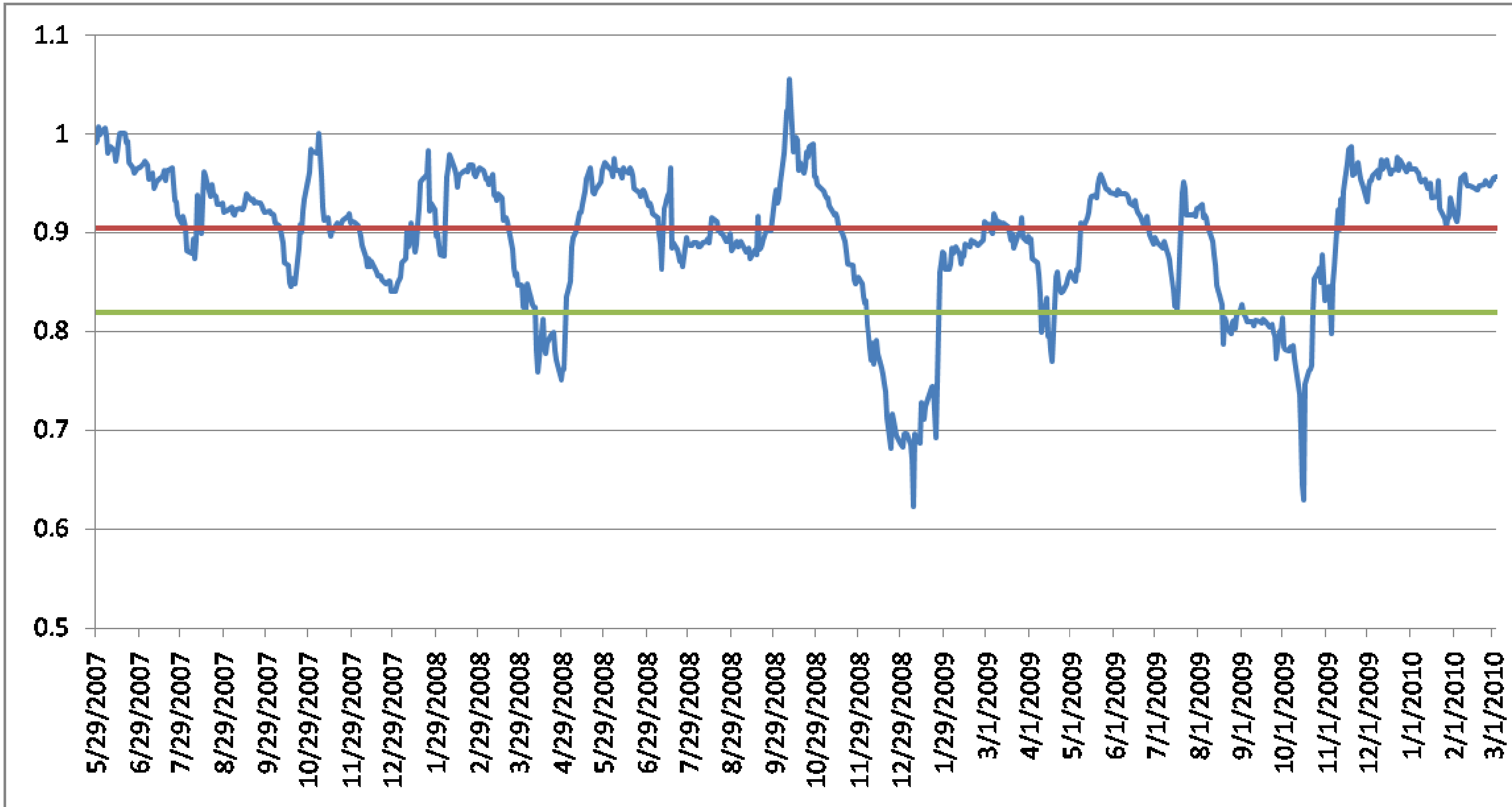
Included observations: 755

p = 1 - q = 0 - constant - manual selection

	Coefficient	Std. Error	t-Statistic	Prob.
c	4.627528	0.03	168.4630632	0
AR(1)	0.98229241	0.01	143.6624447	0

R-squared	0.964800	Mean dependent var	4.628068
Adjusted R-squared	0.964753	S.D. dependent var	0.071188
S.E. of regression	0.013365	Akaike info criterion	-5.782052
Sum squared resid	0.134501	Schwarz criterion	-5.769796
Log likelihood	2184.724802	Durbin-Watson stat	2.225006

AR(1) coefficient for y estimated over a 60-day period



Red= upper bound for MR in 10 days, Green= upper bd for MR in 5 days

Dickey-Fuller over Sep 2008/March 2009

Augmented Dickey-Fuller test statistic			-2.593218	0.284178
Test critical values:	1% level		-4.027516	
	5% level		-3.443485	
	10% level		-3.146482	
Variable	Coefficient	Std. Error	t-Statistic	Prob
tseries(-1)	-0.113728	0.043856	-2.593218	0.010671
D(tseries(-1))	-0.111532	0.090621	-1.230747	0.220785
D(tseries(-2))	0.162647	0.087448	1.859935	0.065303
D(tseries(-3))	0.040018	0.088750	0.450911	0.652854
D(tseries(-4))	-0.267631	0.085738	-3.121501	0.002247
D(tseries(-5))	0.076574	0.086639	0.883828	0.378528
D(tseries(-6))	-0.139433	0.085911	-1.623007	0.107169
D(tseries(-7))	-0.242743	0.082689	-2.935598	0.003980
D(tseries(-8))	0.090026	0.085786	1.049428	0.296056
D(tseries(-9))	0.189077	0.084225	2.244910	0.026575
D(tseries(-10))	-0.106441	0.084083	-1.265896	0.207962
C	0.511426	0.199365	2.565270	0.011521
@trend	0.000090	0.000044	2.058864	0.041636

AR-1 coefficient for the period Sep 2008/march 2009

timeseries: ebay/xlk

Method: Nonlinear Least Squares (Levenberg-Marquardt)

date: 03-03-10 time: 18:40

Included observations: 145

p = 1 - q = 0 - constant - manual selection

	Coefficient	Std. Error	t-Statistic	Prob.
c	4.55489463	0.013841	329.0952386	0
AR(1)	0.88029659	0.035204	25.00582423	-2.2E-16

R-squared	0.813873	Mean dependent var	4.558974
Adjusted R-squared	0.812571	S.D. dependent var	0.045858
S.E. of regression	0.019853	Akaike info criterion	-4.952615
Sum squared resid	0.056364	Schwarz criterion	-4.911557
Log likelihood	361.064616	Durbin-Watson stat	2.312544

Conclusions

ARCH, GARCH: models for volatility of financial series.

Volatility analysis via ARCH and GARCH lead to exponential moving averages of squared returns.

The advantage of GARCH over a fixed window is that GARCH is endogenous. However, fixed estimation windows for volatilities and correlations or exogenous EWMAs also make sense from a risk-management perspective.

Cointegration of stock prices via pairs is not easy to establish econometrically.

Unit root test: tests for stationarity

ARMA, AR: models for mean-reversion

Mean-reversion & pairs trading

Systematic Approach for looking for mean-reversion in Equities

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

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

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)

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

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

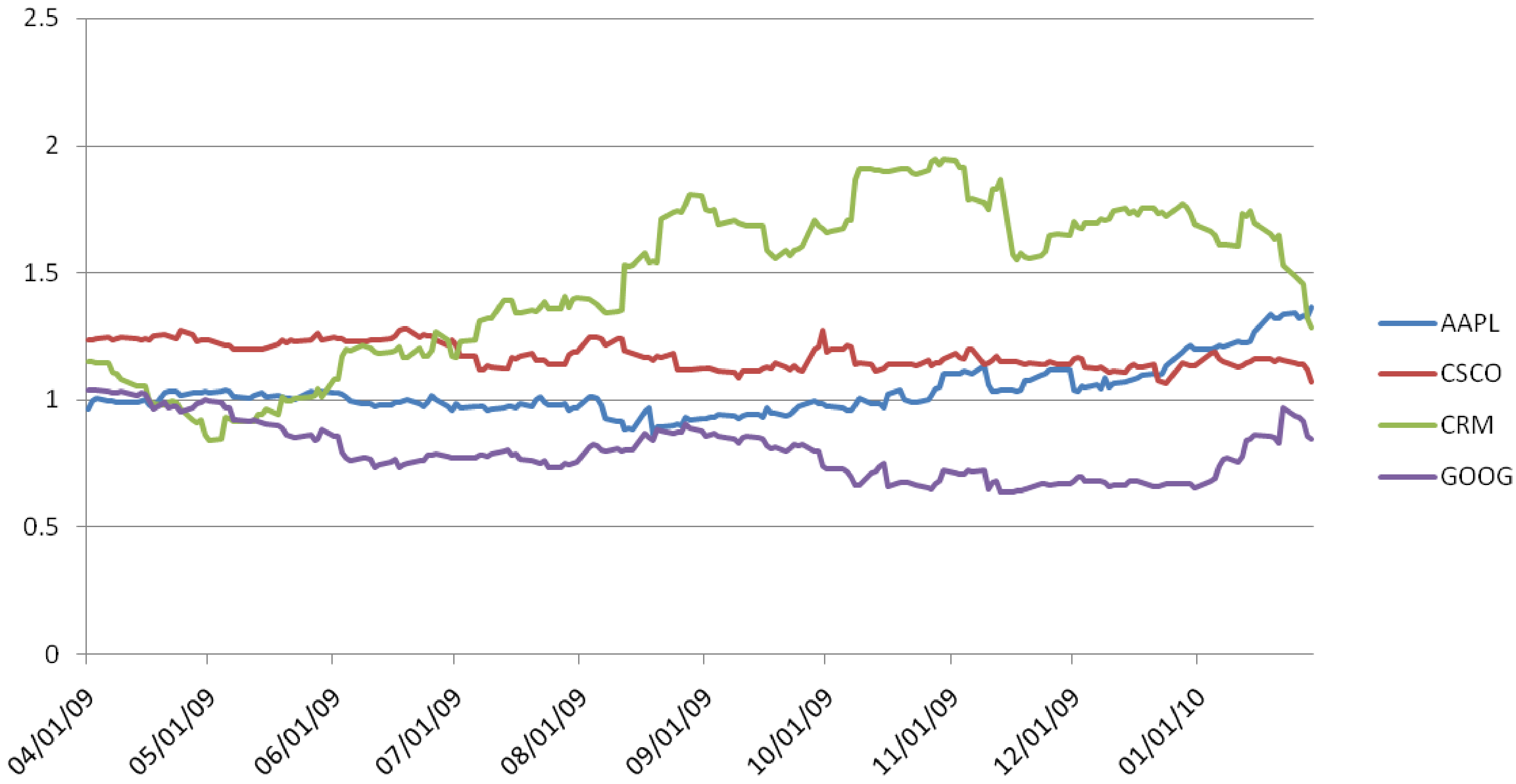
Cross-sectional statistics
for Beta variability

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

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

$X_1, \ldots, X_T$ etf returns

$Y_1, \ldots, Y_T$ stock returns

$w =$ estimation window (in days)

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

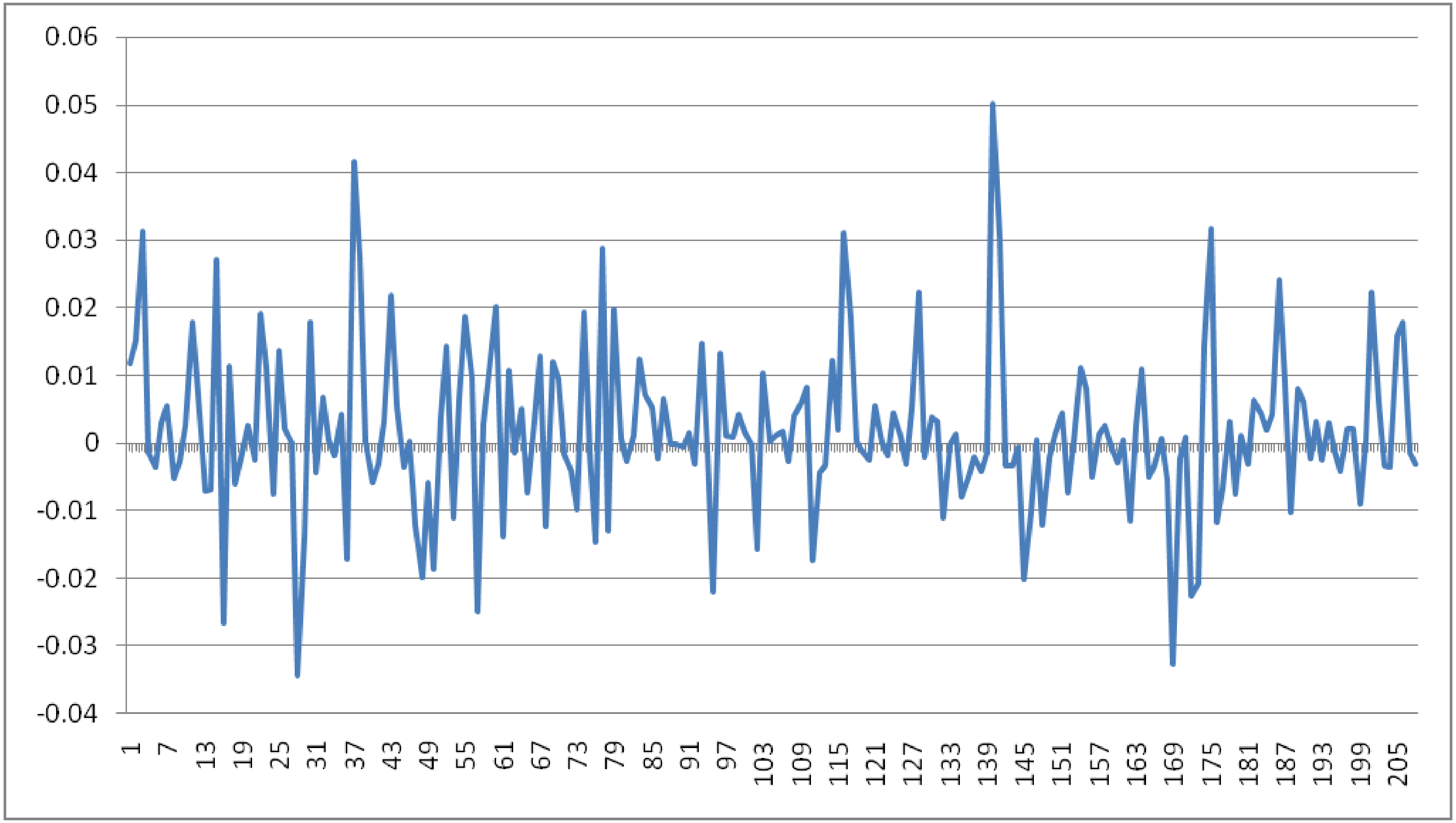
$$\varepsilon_t = Y_t - \beta_{t-w,t} X_t, \qquad t = w+1, w+2, \ldots, T$$

Use window of w days before current date

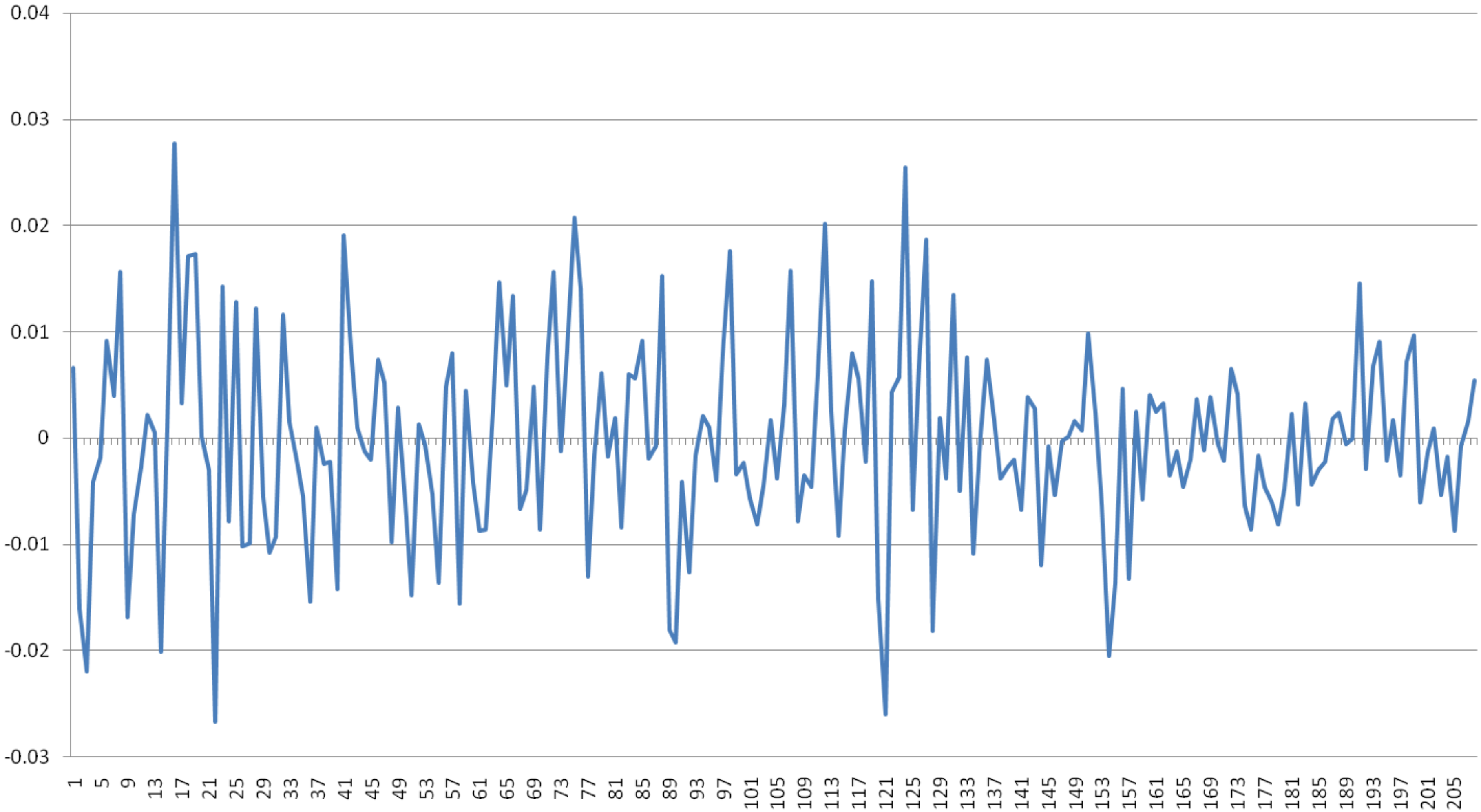
Define $$Z_t := \sum_{k=w+1}^{t} \varepsilon_k$$

``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

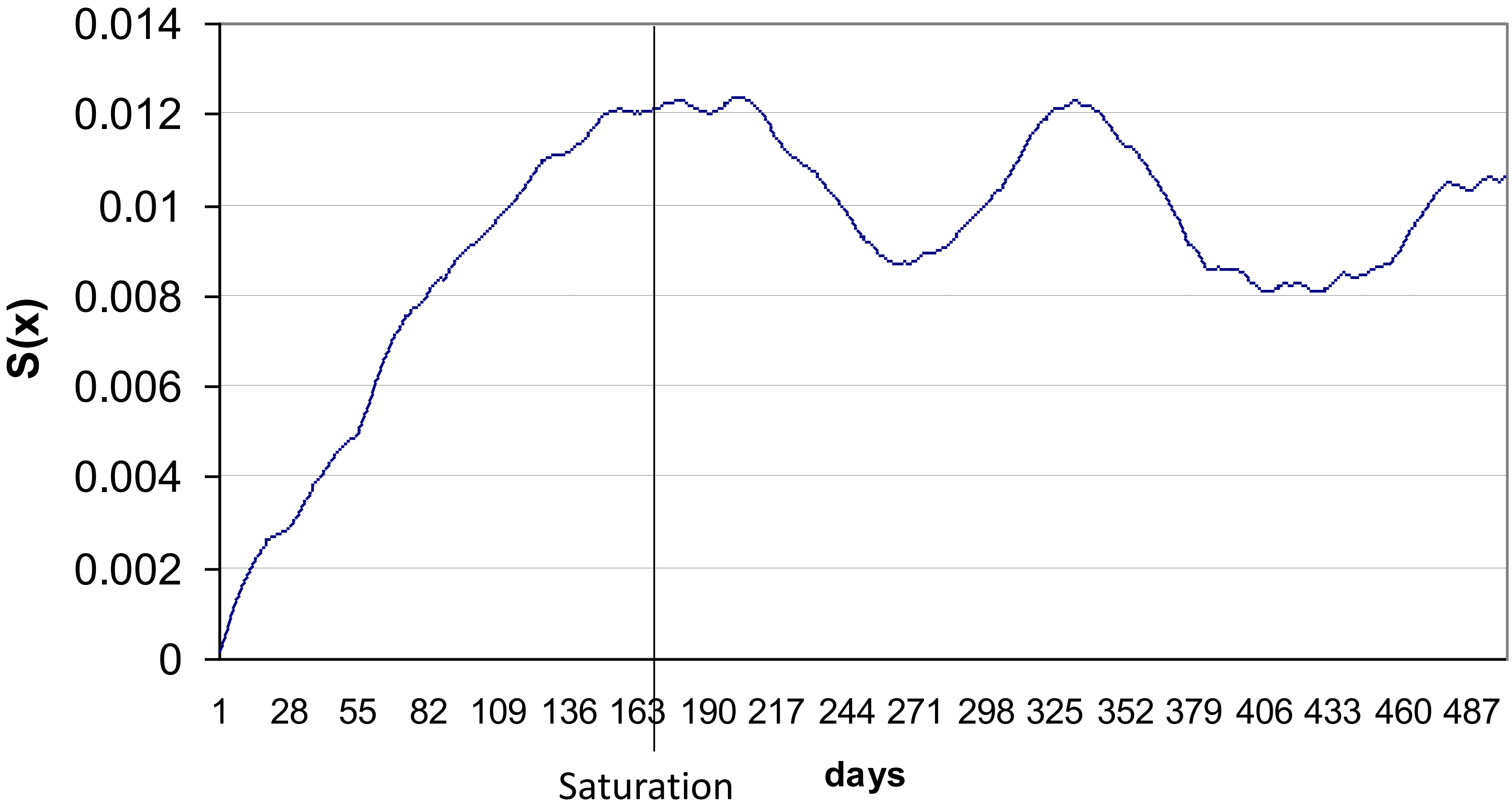
Ticker	CSCO	CTSH	CTXS	DELL	EBAY	EMC	ERTS	FISV	FLIR	GLW
b (slope)	0.93	0.76	1.02	0.93	0.8	0.82	0.89	0.9	0.92	0.97
kappa	17.3	68	-5.7	17.4	56.1	49.4	28.1	27.3	21.6	7.4
tao(in days)	14.6	3.71	-44.2	14.5	4.49	5.1	8.98	9.22	11.7	34.1

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

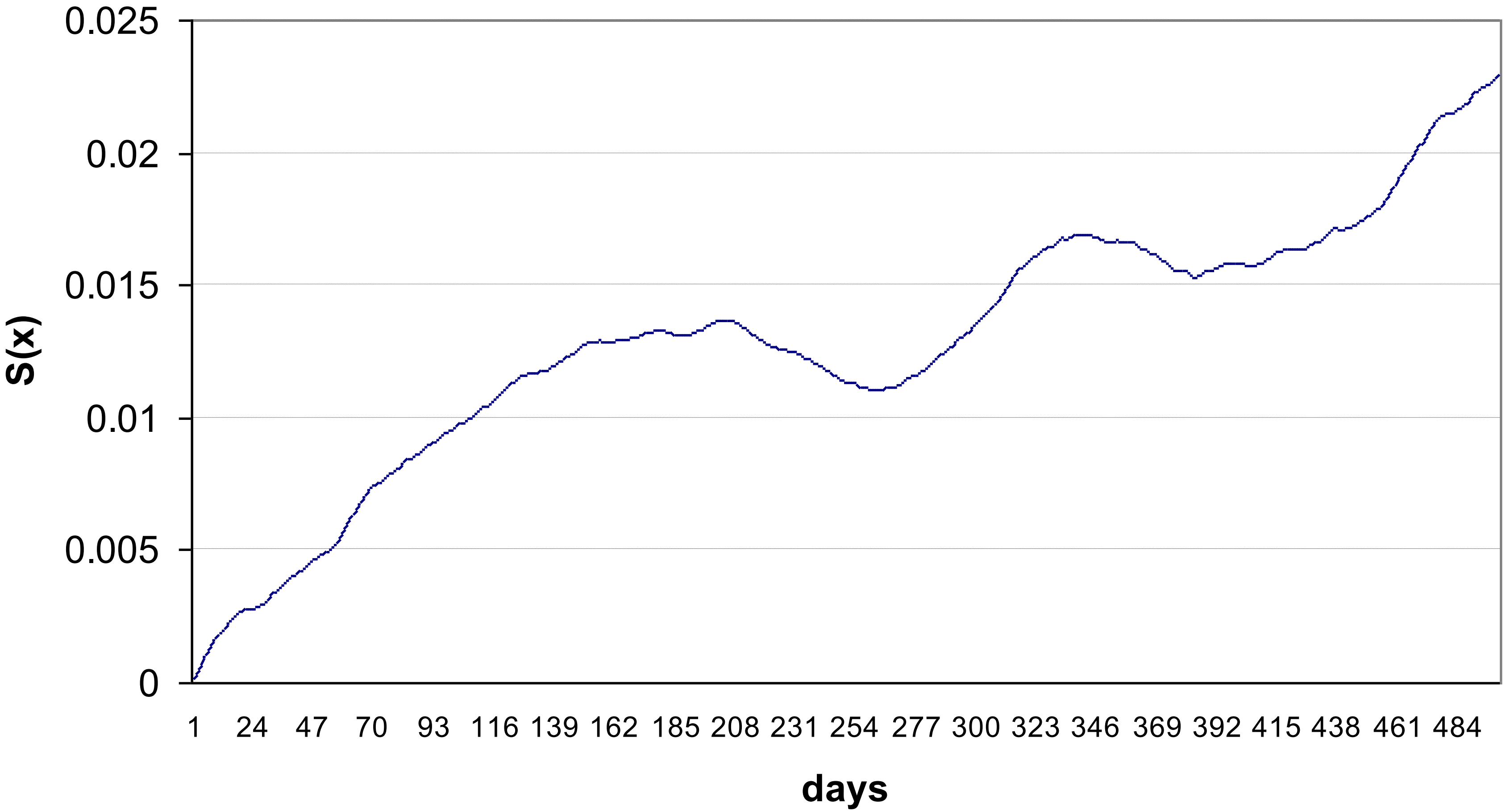
Ticker	NOVL	NTAP	ORCL	QCOM	RHT	SRCL	SYMC	YHOO
b (slope)	0.98	0.9	0.96	0.78	0.88	0.91	0.88	0.91
kappa	5.48	25.8	9.28	61.7	33.4	22.9	31.2	25
tao(in days)	46	9.78	27.2	4.08	7.55	11	8.08	10.1

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



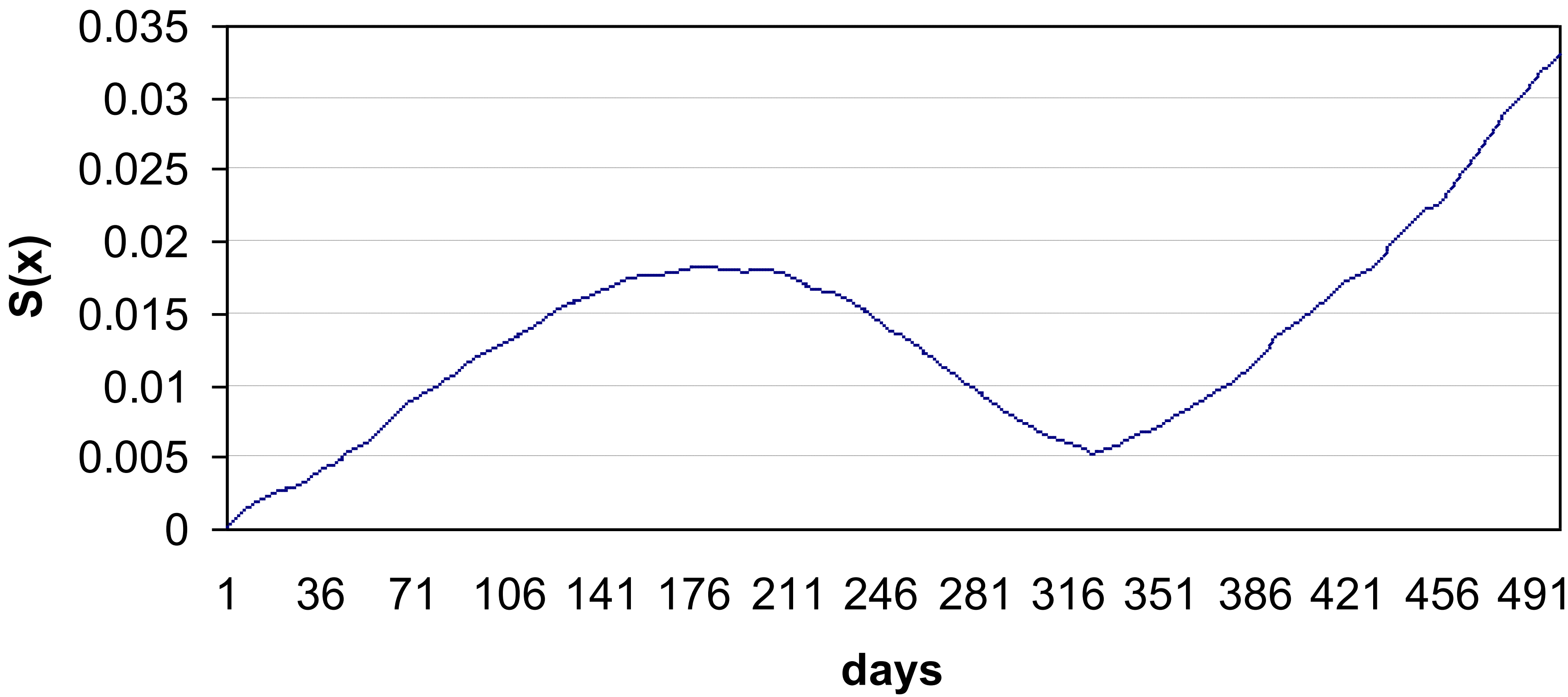
OIH: Oil Services ETF, SLB: Schlumberger-Doll Research

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



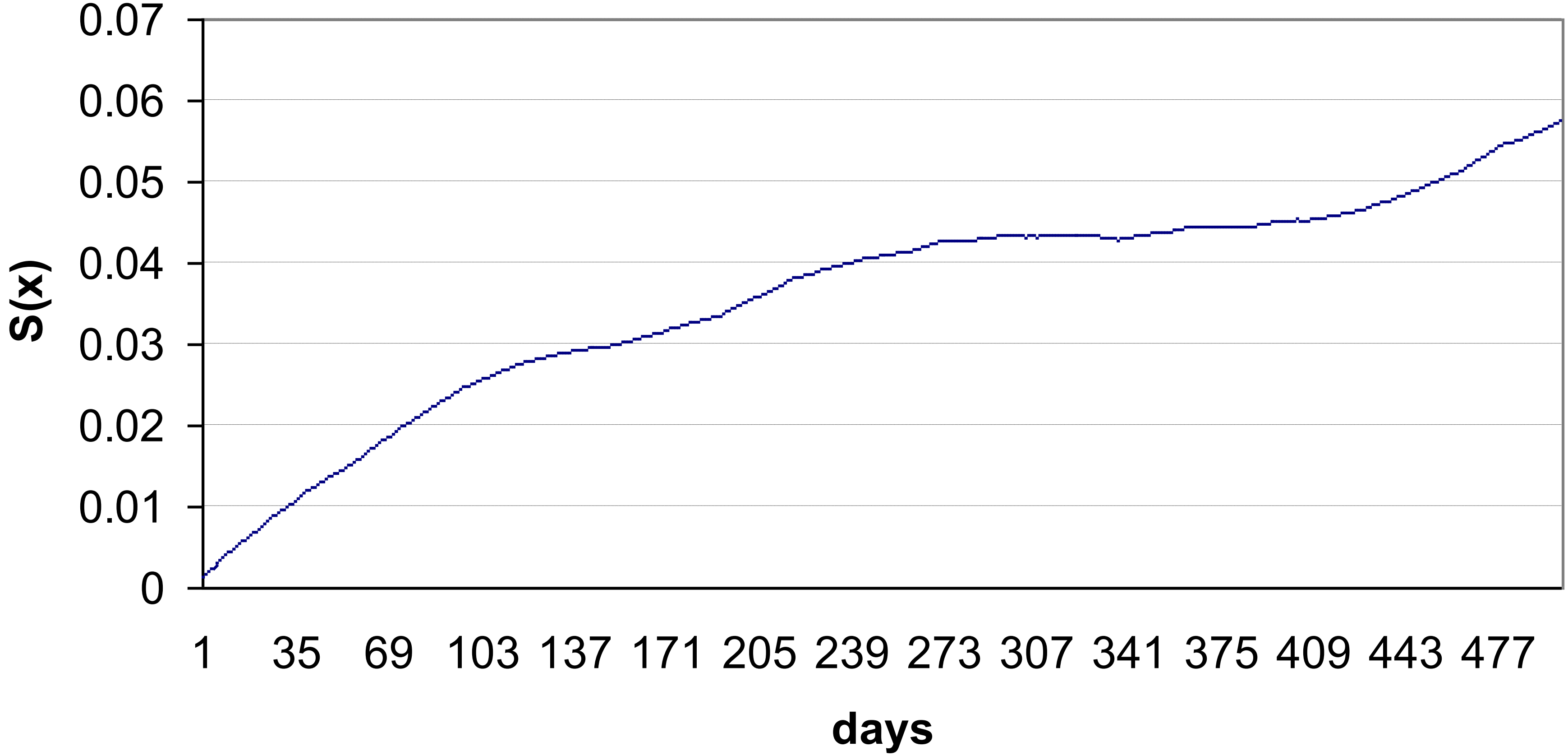
$$P_{n+1} = P_n \times \left(1 + R_{\text{slb}} - R_{\text{oih}}\right), \quad X_n = \ln P_n$$

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

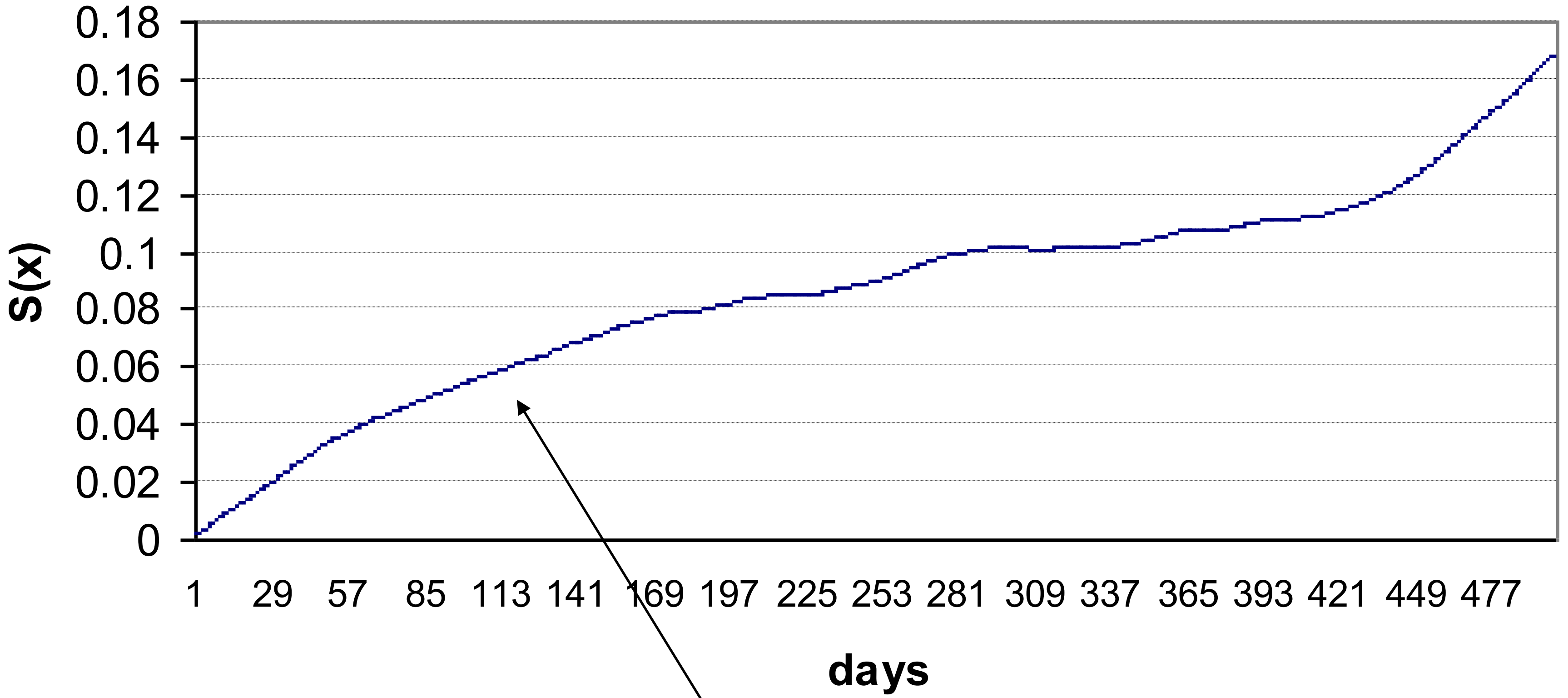


$$P_{n+1} = P_n \times \left(1 + R_{\mathrm{slb}} - \beta_{60\mathrm{d}} \cdot R_{\mathrm{oih}}\right), \quad X_n = \ln P_n$$

Structure Function log (GENZ/IBB)



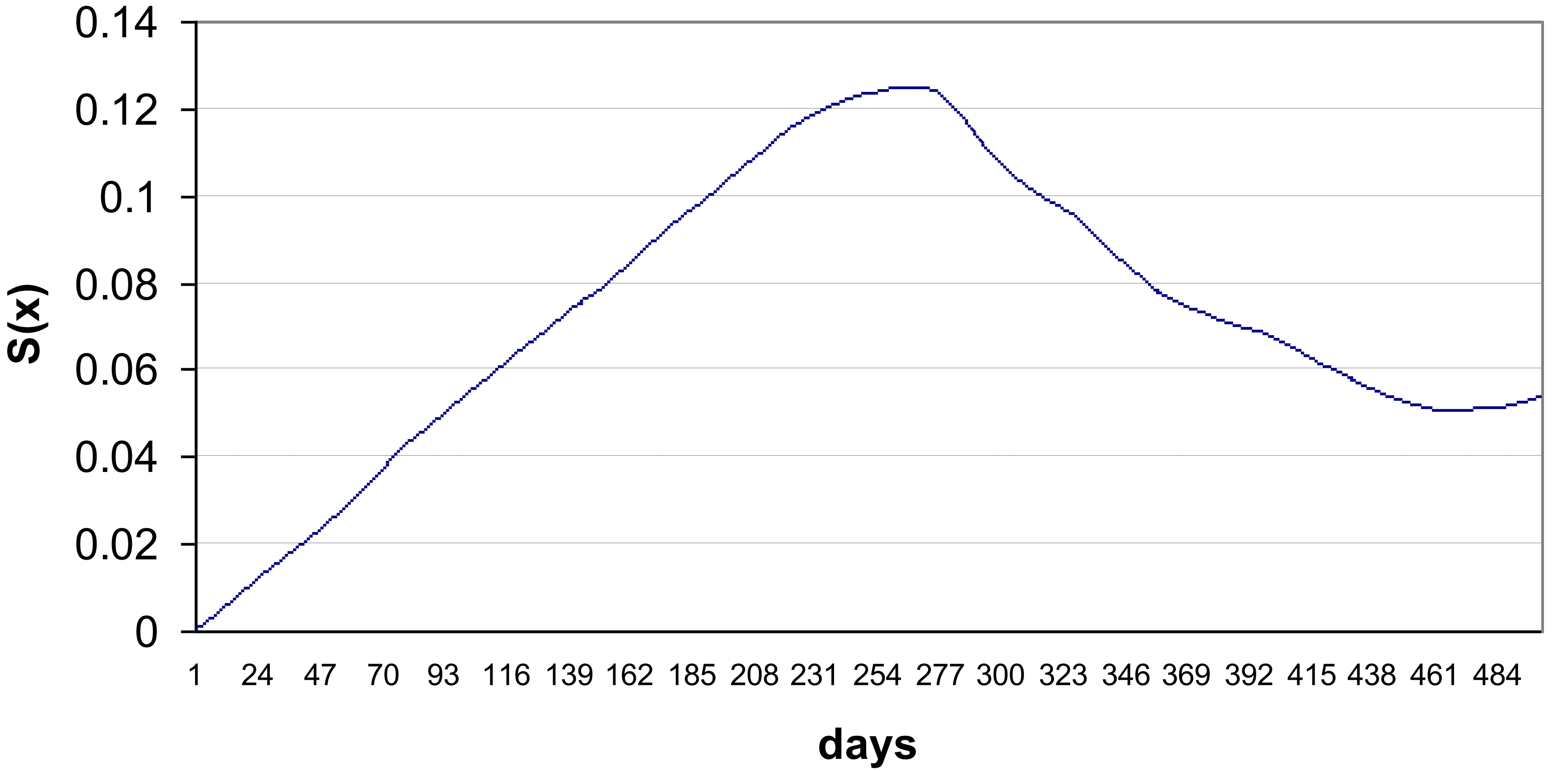
Structure function ln (DNA/GENZ)



Mean-reversion: large negative curvature here.

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