A low-angle, upward-looking photograph of several modern skyscrapers reaching towards a blue sky with scattered white clouds. A commercial airplane is visible in the sky, flying between the buildings. The perspective creates a sense of height and scale.

Chapter 17

Structural Breaks

01.

Motivation

Motivation

- **배경**
 - ML-Based 투자자들은 리스크-조정 수익률이 양의 값일 때 베팅할 것이다
 - 리스크-조정 수익률은 Factor로 판단한다.
 - **Structural Break**는 리스크 조정 수익 최대화를 위해 좋은 피쳐
 - Ex) 평균 회귀 에서 모멘텀 패턴으로 시장 변화
 - **평균 회귀**
 - Buy low, sell high
 - **모멘텀**
 - Higher, Lower
 - 대부분의 시장 참여자들은 이러한 변화를 아주 늦게 발견하고 큰 손해를 봄
- **목적**
 - 본 장에서는 구조적 변화를 측정하는 기법을 알아봄
 - 이 기법을 통해 정보성 특징을 구축할 수 있도록 한다.

02.

Types of structural break tests

Types of structural break tests

- 배경
 - Martingale
 - $E[M(t)|F(s)] = M(s)$ for $0 \leq s \leq t \leq T$
 - Fairness
 - 확률변수의 미래의 기대값이 현재의 확률변수 값과 같음
 - Markov
 - $P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = P(X_{t+1} = i_t | X_t = i_t)$
 - Memoryless
 - 미래 기대값이 과거 경로에 무관하게 계산 시점의 값에만 의존

Types of structural break tests

- 종류
 - CUSUM Test
 - 누적 예측 오차가 white noise로부터 현저히 벗어나는지 테스트
 - 폭발성 Test
 - 백색잡음으로부터의 이탈을 넘어서서 프로세스가 기하급수적 성장이나 붕괴를 보이는지 테스트
 - Random Walk나 Stationary Process와 일치하지 않고, 장기적으로 지속 불가능
 - Right-tail Unit-root Tests
 - 자기 회귀 가정 하에서 기하급수적 성장이나 붕괴가 있는지 평가
 - Sub/Super Martingale Tests
 - 다양한 함수 형태 하에 기하급수적 성장이나 붕괴가 있는지 평가

03.

CUSUM Test

CUSUM Test

- **복습**
 - 2장 CUSUM 필터
 - 바의 이벤트-기반 추출에서 사용
 - 누적 예측 오차가 사정에 정한 임계값을 초과할 경우 바를 추출
- **정의**
 - 자료가 관측될 때마다 반복적으로 회귀 모형을 적합
 - 적합에서 얻은 잔차들의 누적합이 정해진 경계함수를 벗어나는 순간 값이 변했다고 판단
- **분류**
 - 재귀적 잔차에 브라운-더빈-에반스 CUSUM 테스트
 - 레벨에 대한 추-스틴치콤-화이트 CUSUM 테스트

CUSUM Test

- 재귀적 잔차에 대한 브라운-더빈-에반스 CUSUM 테스트
 - 모든 관측값에서 값 y_t 의 전조가 되는 특징 x_t 의 배열 확인
 - 회귀방정식 반복잔차들의 변화추이를 이용하여 회귀계수의 체계적인 움직임 감시
 - X_t
 - 특징 $t \leq T$, x_t 의 시계열로 구성
 - β 의 재귀적 최소 자승 측정
 - $y_t = \beta_t' x_t + \epsilon_t$
 - 부표본 $[1, k+1], [1, k+2], \dots [1, T]$ 에 대한 적합화를 통해 $T-k$ 최소자승($\hat{\beta}_{k+1}, \dots, \hat{\beta}_T$) 추정 생성
 - 표준화된 1단계 이전 재귀적 잔차 $\hat{w}_t = \frac{y_t - \hat{\beta}_{t-1}' x_t}{\sqrt{f_t}} = \frac{(\beta_{t-1} - \hat{\beta}_{t-1})' x_t}{\sqrt{f_t}} = \frac{\epsilon_t}{\sqrt{f_t}}$
 - $f_t = \hat{\sigma}_\epsilon^2 [1 + x_t'(X_t' X_t)^{-1} x_t]$
 - CUSUM 통계량
 - $S_t = \sum \frac{\hat{w}_j}{\hat{\sigma}_w}$
 - $\hat{\sigma}_w^2 = \frac{1}{T-k} \sum (\hat{w}_t - E[\hat{w}_t])^2$
 - $H_0 : \beta_t = \beta \rightarrow S_t \sim N[0, t - k - 1]$

CUSUM Test

- 레벨에 대한 추-스틴치콤-화이트 CUSUM 테스트
 - 이전 기법에서는 x_t 를 퍼지해 단순화하고 $H_0: \beta_t = 0$ 가정
 - 변화가 없다는 것($E_{t-1}[\Delta y_t] = 0$) 예측
 - 로그 가격 y_t 의 로그 가격 y_n 으로부터 표준화된 이탈 계산($t > n$)
- 수식
 - $S_{n,t} = (y_t - y_n)(\hat{\sigma}_t \sqrt{t - n})^{-1}$
 - $\hat{\sigma}_t^2 = (t - 1)^{-1} \sum (\Delta y_i)^2$
- 검정통계량
 - 귀무가설 : $H_0: \beta_t = 0 \rightarrow S_{n,t} \sim N[0,1]$
 - $c_\alpha[n, t] = \sqrt{b_\alpha + \log [t - n]}$

04.

Explosiveness Tests

폭발성 테스트

- 정의
 - 거품
 - 반등과 폭락 모두 의미
- 분류
 - 하나의 거품 테스트
 - 다수의 거품 테스트
 - 거품-폭발-거품 사이클이 계열 정상성 단일-거품 테스트처럼 보이게 한다는 관점에서 보다 견고
- 종류
 - 초-형태 디키-풀러 테스트
 - 상한 증강 디키-풀러

폭발성 테스트

- 배경
 - ADF
 - 수식
 - $y_t = \mu_x + \beta y_{t-1} + \sum \phi \Delta y_{t-i} + \epsilon_t$
 - $H_0: \beta = 1$
 - $H_1: \beta > 1$ (일반적인 경우 $\beta < 1$)
 - 해석
 - 시계열 데이터가 정상성 지니는 지 확인
 - 정상성을 가지지 않는다면 차분을 통해 평균을 일정하게 만들어줘야 함
 - 평균에 대해서만 일어나기 때문에 분산에 대해서는 고려하지 않음
 - 별도의 테스트 필요

폭발성 테스트

- 초-형태 디키-풀러 테스트(Chow-Type Dickey-Fuller Test)
 - 배경
 - 1차 자기 회귀 과정
 - $y_t = \rho y_{t-1} + \epsilon_t$
 - 귀무가설 : y_t 가 랜덤 경로 따름 ($H_0: \rho = 1$)
 - 대립가설 : 랜덤 경로로 시작하지만 시간 τT^* 에 따라 폭발성 프로세스로 변환($\tau^* \in (0,1)$)
 - $H_1: y_t = \begin{cases} y_{t-1} + \epsilon_t & \text{for } t = 1, \dots, \tau^* T \\ \rho y_{t-1} + \epsilon_t & \text{for } t = \tau^* T + 1 \dots, T, \text{ with } \rho > 1 \end{cases}$
 - 시각 T에서 스위치(랜덤 프로세스에서 폭발 프로세스로)가 $\tau^* T$ 에서 발생했는지 테스트
 - $\Delta y_t = \delta y_{t-1} D_t[\tau^*] + \epsilon_t$
 - $D_t[\tau^*]$: 더미변수
 - 0 if $t < \tau^* T$
 - 1 if $t \geq \tau^* T$
 - 귀무가설 ($H_0 : \delta = 0$)
 - 대립가설 ($H_1 : \delta > 1$)
 - $DFC_{\tau^*} = \frac{\widehat{\delta}}{\widehat{\sigma_\delta}}$

폭발성 테스트

- 초-형태 디키-풀러 테스트(Chow-Type Dickey-Fuller Test)
- 단점
 - τ^* 을 알 수 없음
 - $\tau^* \in [\tau_0, 1 - \tau_0]$ 구간 내의 모든 가능한 τ^* 에 대해 테스트
 - $SDFC = \sup\{DFC_{\tau}^*\}$
 - 오직 하나의 발생 날짜 τ^*T 가 있어서 거품이 표본 마지막까지 지속됨(돌아가지 않음)

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)

- 배경

- 표준 단위-루트와 공적분 테스트는 정상성 프로세스와 주기적 붕괴 거품 모델을 효과적으로 구분하지 못함
- 주기적으로 붕괴하는 패턴은 잠재적으로 폭발하는 프로세스보다는 단위-루트나 정상성 자기 회귀로부터 생성된 데이터처럼 보임

- 수식

- $\Delta y_t = \alpha + \beta y_{t-1} + \sum \gamma_l \Delta y_{t-1} + \epsilon_t$
- 귀무가설 : $H_0: \beta \leq 0$
- 대립가설 : $H_1: \beta > 0$

- 해설

- 기준 시점에서 관찰 영역을 한 시점씩 증가하면서 마지막까지 반복적인 ADF 실시

- $SADF_t = \sup\{ADF_{t_0,t}\} = \sup\{\frac{\hat{\beta}_{t_0,t}}{\hat{\sigma}_{\beta_{t_0,t}}}\}$
 - $\hat{\beta}_{t_0,t}$: t_0 에서 시작해 t 에서 끝나는 표본에서 계산
 - τ : 분석에 사용되는 최소 표본 길이
 - t_0 : backward expanding window의 왼쪽 경계
 - $t : \tau, \dots, T$

폭발성 테스트

- $SADF_t$ 와 $SDFC$ 의 차이
 - $SADF_t$
 - 범위
 - $t \in [\tau, T]$
 - 더미 변수를 쓰는 대신 재귀적으로 표본의 시작 확장
 - $t_0 \in [1, t - \tau]$
 - $SDFC$
 - 범위
 - T
 - 중첩된 이중 루프를 (t_0, t) 의 모든 조합에 대해 시행
 - 레짐 스위치나 발생 날짜 가정하지 않음

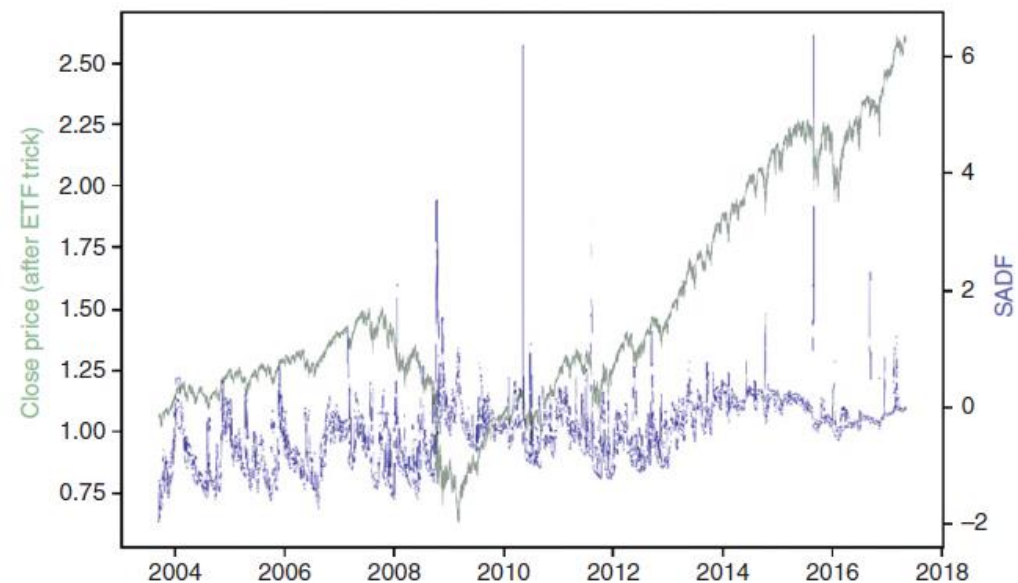


FIGURE 17.1 Prices (left y-axis) and SADF (right y-axis) over time

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)
 - 원시 가격 대 로그 가격
 - 배경
 - 논문에서 보면 구조적 변화 테스트를 원시 가격에 수행한 것을 흔히 볼 수 있음
 - 하지만 원시 가격보다는 로그 가격이 선호되어야 함
 - 거품과 폭발을 포함한 장기간의 시계열의 경우 로그 가격이 필수
 - 해설
 - 원시 가격으로 작업
 - 원시 가격에 대해서 ADF가 기각 \rightarrow 가격이 유한한 분산을 가진 정상성 $\rightarrow r_t = \frac{y_t}{y_{t-1}} - 1$ 이 시간 불변성이 아님
 - 수익률의 변동성은 가격 분산을 일정하게 유지하기 위해 가격이 오르면 내려가고 내리면 올라감
 - ADF를 원시 가격에 수행할 때는 수익률의 분산이 가격 수준에 대해 불변성이 아니라고 가정
 - 수익률 분산은 가격 수준에 불변적이면 모델은 구조적으로 이분산성
 - 로그 가격으로 작업
 - $ADF : \Delta \log[y_t] \propto \log[y_{t-1}]$
 - 변수에 변화를 줄 경우 $x_t = k y_t$
 - $\log[x_t] = \log[k] + \log[y_t]$
 - $ADF : \Delta \log[x_t] \propto \log[x_{t-1}] \propto \log[y_{t-1}]$
 - 로그 가격에 기초한 이 대체 식에서 가격 레벨은 수익률의 평균에는 영향을 미치지만 수익률의 변동성에는 영향 미치지 못함
 - $k \neq 1$ 일 경우 큰 차이가 발생할 수 있음

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)
 - 계산 복잡도
 - $O(n^2)$
 - 전체 표본 길이 T 에 대해 SADF가 요구하는 ADF 검정 개수가 다음과 같기 때문
 - $\sum t - \tau + 1 = \frac{1}{2}(T - \tau + 2)(T - \tau + 1) = \binom{T - \tau + 2}{2}$
 - 단일 ADF 회귀를 해결하려면 표에 나열한 실수 연산 횟수만큼 계산 필요
 - 매번 ADF를 계산할 때 총 $f(N, T) = N^3 + N^2(2T + 3) + N(4T - 1) + 2T + 2$ FLOP 필요
 - T 가 커질수록 기하급수적으로 증가함
 - 이중 루프가 아닌 병렬처리할 필요가 있음(20장 고성능 컴퓨팅)

행렬 연산	FLOPs
$o_1 = X'y$	$(2T - 1)N$
$o_1 = X'X$	$(2T - 1)N^2$
$o_3 = o_2^{-1}$	$N^3 + N^2 + N$
$o_4 = o_3 o_1$	$2N^2 - N$
$o_5 = y - X o_4$	$T + (2N - 1)T$
$o_6 = o_5' o_5$	$2T - 1$
$o_7 = o_3 o_6 \frac{1}{T - N}$	$2 + N^2$
$o_8 = \frac{o_{4[0,0]}}{\sqrt{o_{7[0,0]}}}$	1

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)

- 지수적 행동 조건

- 로그 가격의 제로-랙 식

- $\Delta \log[y_t] = \alpha + \beta \log[y_{t-1}] + \epsilon_t$

- $\log[\tilde{y}_t] = (1 + \beta) \log[\tilde{y}_{t-1}] + \epsilon_t$

- $\log[\tilde{y}_t] = \log[y_t] + \frac{\alpha}{\beta}$

- $E[\log[\tilde{y}_t]] = (1 + \beta)^t \log[\tilde{y}_0], \quad E[\log[y_t]] = -\frac{\alpha}{\beta} + (1 + \beta)^t (\log[y_0] + \frac{\alpha}{\beta})$

- 상태 조건

- 안정성

- $\beta < 0 \rightarrow \lim_{t \rightarrow \infty} E[\log[y_t]] = -\frac{\alpha}{\beta}$

- 불안정은 $\log[y_t] - \left(-\frac{\alpha}{\beta}\right) = \log[\tilde{y}_t]$

- $t = -\frac{\log[2]}{\log[1+\beta]} (\text{반감기})$ 에서 $\frac{E[\log[\tilde{y}_t]]}{\log[\tilde{y}_0]} = (1 + \beta)^t = \frac{1}{2}$

- 단위·루트

- $\beta = 0$

- 시스템은 비정상성이고, 마틴게일로 행동

- 폭발성

- $\beta > 0$

- $\lim_{t \rightarrow \infty} E[\log[y_t]] = \begin{cases} -\infty & \text{if } \log[y_0] < \frac{\alpha}{\beta} \\ +\infty & \text{if } \log[y_0] > \frac{\alpha}{\beta} \end{cases}$

폭발성 테스트

- 로그 가격의 제로-랙 식

- $\Delta \log [y_t] = \alpha + \beta \log [y_{t-1}] + \epsilon_t$

- $\log [\tilde{y}_t] = (1 + \beta) \log [\tilde{y}_{t-1}] + \epsilon_t$

- $\log [\tilde{y}_t] = \log [y_t] + \frac{\alpha}{\beta}$

- $E[\log [\tilde{y}_t]] = (1 + \beta)^t \log [\tilde{y}_0], \quad E[\log [y_t]] = -\frac{\alpha}{\beta} +$

$$(1 + \beta)^t (\log [y_0] + \frac{\alpha}{\beta})$$

- 상태 조건

- 안정성

- $\beta < 0 \rightarrow \lim_{t \rightarrow \infty} E[\log [y_t]] = -\frac{\alpha}{\beta}$

- 불안정은 $\log [y_t] - \left(-\frac{\alpha}{\beta}\right) = \log [\tilde{y}_t]$

- $t = -\frac{\log [2]}{\log [1 + \beta]} (\text{반감기})$ 에서 $\frac{E[\log [\tilde{y}_t]]}{\log [\tilde{y}_0]} = (1 + \beta)^t = \frac{1}{2}$

- 단위-루트

- $\beta = 0$

- 시스템은 비정상성이고, 마틴게일로 행동

- 폭발성

- $\beta > 0$

- $\lim_{t \rightarrow \infty} E[\log [y_t]] = \begin{cases} -\infty & \text{if } \log [y_0] < \frac{\alpha}{\beta} \\ +\infty & \text{if } \log [y_0] > \frac{\alpha}{\beta} \end{cases}$

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)
 - 분위 수 ADF
 - SADF는 t 값에서 계열 상한 취함
 - $SADF_t = \sup_{t_0 \in [1, t-\tau]} \{ADF_{t_0}, t\}$
 - 극단값을 선택하면 안정성 문제가 야기돼 SADF 계산이 표본 빈도와 표본의 특정 타임스탬프에 종속돼 상당히 변할 수 있음
 - 극단값보다 안정적인 계산
 - $s_t = \{ADF_{t_0, t}\}_{t_0 \in \{0, t_1 - \tau\}}$ 라 가정
 - s_t 의 q 분위수 $Q_{t, q} = Q[s_t, q]$ 를 집중성의 척도로 정의
 - $\dot{Q}_{t, q, v} = Q_{t, q+v} - Q_{t, q-v}, 0 < v \leq \min\{q, 1 - q\}$ 를 고ADF값의 분산성 척도로 정의

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)
 - 조건부 ADF
 - 조건부 모멘트를 계산함으로써 안정성 문제 해결
 - 전개
 - $f[x]$ 는 $s_t = \{ADF_{t_0,t}\}_{t_0 \in [0, t_1 - \tau]}, x \in s_t$ 의 확률분포 함수
 - $C_{t,q} = K^{-1} \int_{Q_{t,q}}^{\infty} x f[x] dx$: ADF값의 집중성 척도
 - $C_{t,q} \leq SADF_t$
 - $c_{t,q} = \sqrt{K^{-1} \int_{Q_{t,q}}^{\infty} (x - C_{t,q})^2 f[x] dx}$: 고ADF값의 분산성 척도
 - $K = \int_{Q_{t,q}}^{\infty} f[x] dx$
 - 결과
 - $\frac{SADF_t - C_{t,q}}{C_{t,q}}$ 는 $SADF_t$ 가 이상값에 민감하기 때문에 $C_{t,q}$ 가 작더라도 큰 값에 도달할 수 있음

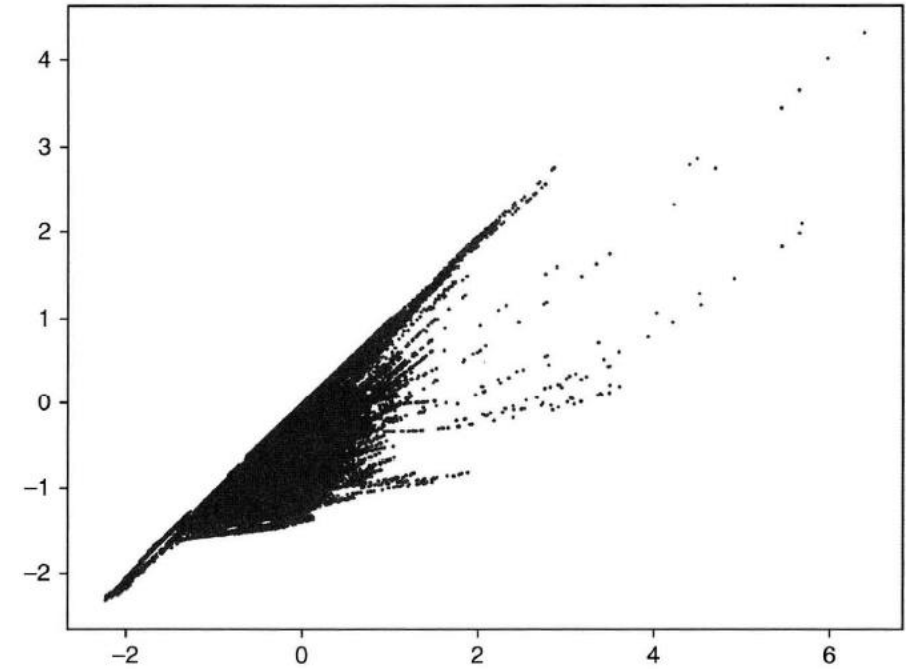


그림 17-2 SADF(x축) 대 CADF(y축)

폭발성 테스트

- 상한 증강 디키-풀러(Supremum Augmented Dickey-Fuller)
 - 조건부 ADF
 - s_t 의; 이상값이 $SADF_t$ 를 상향으로 편향하는 경향을 보임

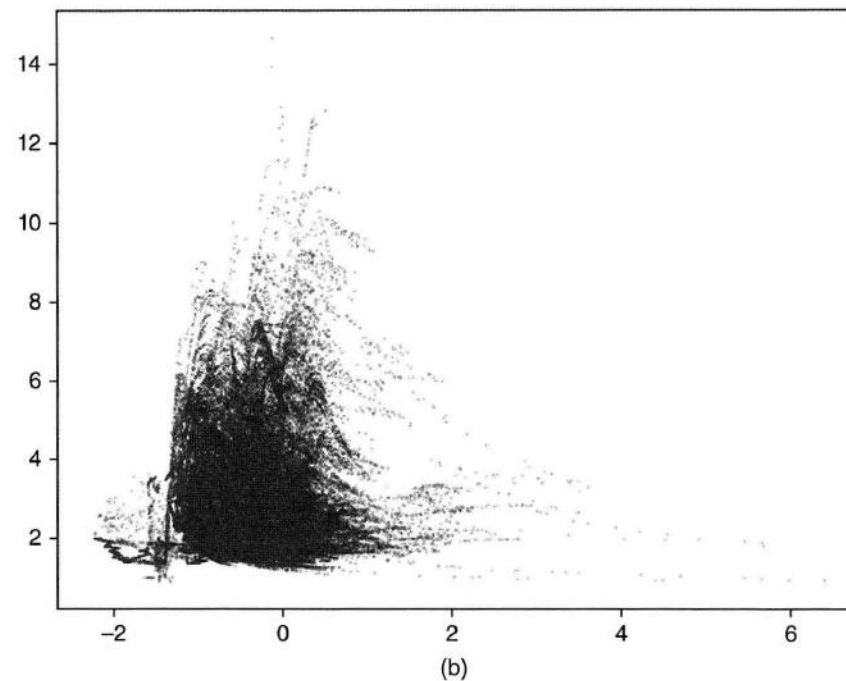
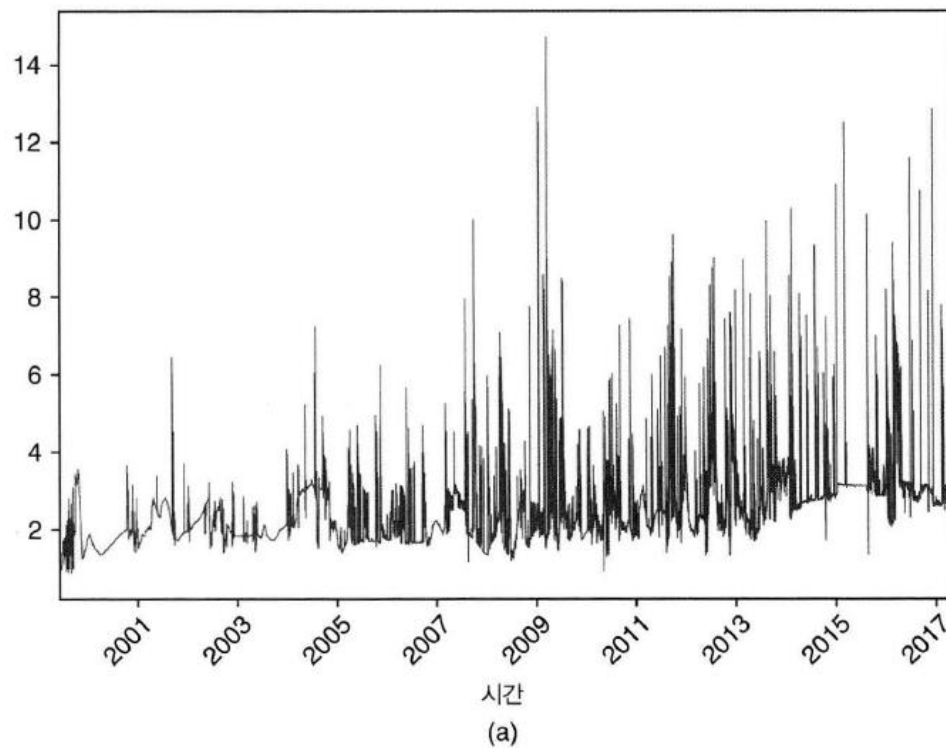


그림 17-3(a) 시간에 대한 $(SADF_t - C_{t,a})/\hat{C}_{t,a}$ (b) $(SADF_t - C_{t,a})/\hat{C}_{t,a}$ (y축), $SADF_t$ (x축)의 함수로 표시

폭발성 테스트

- 서브 또는 슈퍼-마팅게일 검정
 - 특징
 - ADF식에 의존하지 않는 폭발성 테스트
 - 목적 : 관측값 y_t 에 대해 폭발성 시간 경향 존재 테스트
 - $H_0: \beta = 0, H_1: \beta \neq 0$
 - 종류
 - 다항 경향(SM-Poly1) : $y_t = \alpha + \gamma t + \beta t^2 + \epsilon_t$
 - 다항 경향(SM-Poly2) : $\log[y_t] = \alpha + \gamma t + \beta t^2 + \epsilon_t$
 - 지수 경향(SM-Expt) : $y_t = \alpha e^{\beta t} + \epsilon_t \rightarrow \log[y_t] = \log[\alpha] + \beta t + \zeta_t$
 - 멱 경향(SM-Power) : $y_t = \alpha t^\beta + \epsilon_t \rightarrow \log[y_t] = \log[\alpha] + \beta \log[t] + \zeta_t$
 - 전개
 - 각 포인트 $t = \tau, \dots, T$
 - $SMT_t = \sup \left\{ \frac{|\hat{\beta}_{t_0,t}|}{\sigma_{\hat{\beta}_{t_0,t}}} \right\}$ 계산 여기에 수식을 입력하십시오. $SMT_t = \sup \left\{ \frac{|\hat{\beta}_{t_0,t}|}{\sigma_{\hat{\beta}_{t_0,t}}(t-t_0)^\varphi} \right\}$
 - 절대값 : 폭발적 성장과 붕괴 모두 관심
 - 약한 장기-지속 거품의 $\hat{\sigma}_\beta^2$ 는 강한 단기-지속 거품보다 작을 수 있으므로 장기-지속 거품쪽으로 편향이 생길 수 있음
 - 큰 표본 길이 고려 필요

A low-angle, upward-looking photograph of several modern skyscrapers reaching towards a bright blue sky with scattered white clouds. A white commercial airplane is visible in the center of the frame, flying upwards. The perspective creates a sense of height and grandeur.

Thank you