

Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values? 精读笔记

蔡巍

June 19, 2025

Contents

1	文章主要内容介绍	1
2	主要内容推导	2
2.1	The rational bubble	2
2.2	ADF test	6
2.3	Econometric analysis of explosive process	9
2.4	Finite sample bias correction via indirect inference	10
2.5	Estimating origination and collapse dates	12
3	数学基础	23
3.1	现金流折现以及现值计算	23
3.2	范数	24
3.3	o_p 和 O_p 符号	24
3.4	Donsker 定理 (Donsker' s Theorem)	25
3.5	伊藤积分 (Itô Integral)	26

1 文章主要内容介绍

问题的提出：1990 年代，纳斯达克指数因互联网和科技股的热潮经历了显著上涨，尤其是在 1995 年至 2000 年间，最终在 2000 年 3 月达到历史高点 5048.62 点，随后大幅下跌，摧毁了大量的股东财富。如何检测金融市场中的爆炸性行为（可能表明泡沫的存在），并确定其发生和结束的时间。

方法：论文提出了一种新的递归测试方法（recursive test procedure），用于检测时间序列中的爆炸性行为，并能够“时间标记”（date stamp）泡沫的起源和崩溃。这种方法基于右尾单位根测试（right-tailed unit root test）和 sup 测试，结合了温和爆炸性过程（mildly explosive processes）的极限理论，克服了传统单位根测试无法有效检测周期性崩溃泡沫（periodically collapsing bubbles）的局限性。递归测试通过逐次增加样本数据，分析时间序列的动态行为，能够捕捉短期的爆炸性特征。

实证结果：应用该方法分析 1973 年至 2005 年的纳斯达克指数数据，研究发现：爆炸性行为始于 1995 年 7 月，早于 Greenspan 1996 年 12 月的“非理性繁荣”言论，表明 Greenspan 的评论有数据支持。爆炸性行为持续到 2000 年 9 月至 2001 年 3 月之间，泡沫在 2000 年 2 月达到顶峰。股息序列没有显示爆炸性行为，表明股价的爆炸性主要由泡沫成分（ b_t ）驱动。

2 主要内容推导

为了更加深刻地了解这个方法，我需要详细推导以下内容：

2.1 The rational bubble

本节旨在从理论角度定义和识别市场中的“非理性繁荣”现象。首先，基于现值理论，我们将股票价格分解为两个组成部分：基本面价格（由未来股息现值决定）和泡沫成分。通过分析发现，泡沫成分具有爆炸性自回归特征，这意味着其增长率会随时间加速增长。为了验证这种爆炸性行为是否确实由泡沫驱动，我们进一步考察了股价和股息的动态特性。在保持折现率不变的条件下，如果观察到股价呈现爆炸性增长而股息保持相对稳定，则可以推断这种爆炸性行为主要来源于泡沫成分的膨胀，而非基本面因素。

推导：理性泡沫从下面这个式子开始：

$$P_t = \frac{1}{1+R} E_t(P_{t+1} + D_{t+1}), \quad (1)$$

迭代：

$$P_t = \frac{1}{1+R} E_t\left(\frac{1}{1+R} E_{t+1}(P_{t+2} + D_{t+2}) + D_{t+1}\right) \quad (2)$$

...

$$P_t = \underbrace{\lim_{n \rightarrow \infty} \frac{1}{1+R}^n E_t(P_{t+n})}_{\text{未来股价现值 } B_t} + \underbrace{\sum_{n=1}^{\infty} \frac{1}{1+R}^n E_t(D_{t+n})}_{\text{未来股利现值 } P_t^f} \quad (3)$$

now we have:

$$P_t = B_t + P_t^f \quad (4)$$

take a log:

$$\log P_t = \log B_t + \log P_t^f \quad (5)$$

we can re-write it as:

$$p_t = b_t + p_t^f \quad (6)$$

from 1 we also have:

$$\log(P_t) = \log\left(\frac{1}{1+R} E_t(P_{t+1} + D_{t+1})\right) = -\log(1+R) + \log(E_t(P_{t+1} + D_{t+1})) \quad (7)$$

now we see the (3) in the paper:

$$p_t^f = \frac{\kappa - \gamma}{1 - \rho} + (1 - \rho) \sum_{i=0}^{\infty} \rho^i E_t d_{t+1+i}, \quad (8)$$

proof:

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1 \quad (9)$$

so:

$$1 + R_t = \frac{P_t + D_t}{P_{t-1}} \quad (10)$$

take a log, and note $\gamma_t = \log(1 + R_t)$:

$$\gamma_t = \log(1 + R_t) = \underbrace{\log(P_t + D_t) - \log(P_{t-1})}_{?} = \log(P_t + D_t) - p_{t-1} \quad (11)$$

so we solve the ? 我觉得这部分证明不一定对但是能得到想要的结果:

先定点: (P,D), 然后在 (P,D) 处, 用泰勒展开¹:

$$P_t = P * e^{\Delta p_t} \Rightarrow \Delta p_t = p_t - p \Rightarrow P_t = P(1 + \Delta p_t) \quad (12)$$

$$D_t = D * e^{\Delta d_t} \Rightarrow \Delta d_t = d_t - d \Rightarrow D_t = D(1 + \Delta d_t) \quad (13)$$

现在我们回来处理 $\log(P_t + D_t)$: ($\frac{P}{P+D} = \rho$)

$$\begin{aligned} \log(P_t + D_t) &= \log(P(1 + \Delta p_t) + D(1 + \Delta d_t)) \\ &= \log(P + P\Delta p_t + D + D\Delta d_t) \\ &= \log\left((P + D)\left(1 + \frac{P\Delta p_t}{P + D} + \frac{D\Delta d_t}{P + D}\right)\right) \\ &= \log(P + D) + \log\left(1 + \rho\Delta p_t + (1 - \rho)\Delta d_t\right) \\ &= \log(P + D) + \rho\Delta p_t + (1 - \rho)\Delta d_t \\ &= \log(P + D) + \rho(p_t - p) + (1 - \rho)(d_t - d) \end{aligned}$$

¹泰勒展开: $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + o(x - x_0)^2$, 特殊的在高中常用的: $e^x \approx 1 + x$, $\ln(1 + x) \approx x$

P 和 D 是定点

$$\begin{aligned}
\gamma_t &= \log(P + D) + \rho(p_t - p) + (1 - \rho)(d_t - d) - p_{t-1} \\
&= \log(P + D) + \rho p_t - \rho p + (1 - \rho)d_t - (1 - \rho)d - p_{t-1} \\
&= \rho p_t + (1 - \rho)d_t - p_{t-1} + \kappa
\end{aligned}$$

其中 κ 就是剩下的几项:

$$\begin{aligned}
\kappa &= \log(P + D) - \rho p - (1 - \rho)d \\
&= \log(P + D) - \frac{P}{P + D} \log P - \frac{D}{P + D} \log D \\
&+ \frac{P}{P + D} \log(P + D) + \frac{D}{P + D} \log(P + D) - \frac{P}{P + D} \log(P + D) - \frac{D}{P + D} \log(P + D) \\
&= \log(P + D) - \frac{P}{P + D} \log\left(\frac{P}{P + D}\right) - \frac{D}{P + D} \log\left(\frac{D}{P + D}\right) - \log(P + D) \\
&= -\rho \log \rho - (1 - \rho) \log(1 - \rho) \\
&= -\rho \log \rho - (1 - \rho) \log\left(\frac{1}{\rho} - 1\right) + \log \rho \\
&= -\rho \log \rho - (1 - \rho) \log\left(\frac{1}{\rho} - 1\right) - (1 - \rho) \log \rho \\
&= -\log \rho - (1 - \rho) \log\left(\frac{1}{\rho} - 1\right)
\end{aligned}$$

这里就化简得和原文一模一样了。

对 γ_t 求期望:

$$\begin{aligned}
E_t(\gamma_t) &= \rho E_t(p_{t+1}) + (1 - \rho) E_t(d_{t+1}) - E_t(p_t) + \kappa \\
&= \rho E_t(p_{t+1}) + (1 - \rho) E_t(d_{t+1}) - p_t + \kappa
\end{aligned}$$

移项, 将 p_t 移到左边:

$$\begin{aligned}
p_t &= \rho E_t(p_{t+t}) + (1 - \rho) E_t(d_{t+1}) + \kappa - \gamma \\
&= \underbrace{\lim_{n \rightarrow \infty} \rho^n E_t(p_{t+n})}_{\rightarrow 0, p_t^f = p_t, b_t} + \sum_{n=0}^{\infty} (1 - \rho)^n E_t(d_{t+n}) + \sum_{i=0}^{\infty} \rho^i \kappa - \sum_{j=0}^{\infty} \rho^j \gamma \\
&= \frac{\kappa - \gamma}{1 - \rho} + \sum_{n=0}^{\infty} \rho^n E_t(d_{t+n})
\end{aligned}$$

换成与原文 notation 一样:

$$b_t = \lim_{i \rightarrow \infty} \rho^i E_t(p_{t+i}) \quad (14)$$

$$\begin{aligned}
b_{t+1} &= \lim_{i \rightarrow \infty} \rho^i E_{t+1}(p_{t+i+1}) \\
&= \lim_{j \rightarrow \infty} \rho^{j-1} E_{t+1}(p_{t+1+(j-1)}) \\
&= \frac{1}{\rho} \lim_{j \rightarrow \infty} \rho^j E_{t+1}(p_{t+j})
\end{aligned}$$

在求 $E_t(b_{t+1})$ 时，使用迭代期望公式 $E_t(E_{t+1}(X)) = E_t(X)$ ：

$$\begin{aligned}
E_t(b_{t+1}) &= E_t\left(\frac{1}{\rho} \lim_{j \rightarrow \infty} \rho^j E_{t+1}(p_{t+j})\right) \\
&= \frac{1}{\rho} E_t\left(\lim_{j \rightarrow \infty} \rho^j E_{t+1}(p_{t+j})\right) \\
&= \frac{1}{\rho} \lim_{j \rightarrow \infty} \rho^j E_t(p_{t+j}) \\
&= \frac{1}{\rho} b_t
\end{aligned}$$

当然， ρ 还可以进一步处理成 $\frac{1}{1+\exp(\overline{d-p})}$ 的形式：

$$\rho = \frac{P}{P+D} = \frac{1}{1+\frac{D}{P}} \quad (15)$$

$$\log\left(\frac{D}{P}\right) = \log(D) - \log(P) = \overline{d-p} \Rightarrow \frac{D}{P} = \exp(\overline{d-p}) \quad (16)$$

so, we get the same equation as (4) in the paper:

$$E_t(b_{t+1}) = \frac{1}{\rho} b_t = \frac{1}{1+\exp(\overline{d-p})} b_t \quad (17)$$

所以写成 (5) 式:

$$b_t = \frac{1}{\rho} b_{t-1} + \varepsilon_{b,t} \equiv (1+g)b_{t-1} + \varepsilon_{b,t}, E_{t-1}(\varepsilon_{b,t}) = 0 \quad (18)$$

$$d_t - p_t = d_t - \frac{\kappa - \gamma}{1 - \rho} - (1 - \rho) \sum_{i=0}^{\infty} \rho^i E_t d_{t+1+i} \quad (19)$$

$\Delta d_{t+1+i} = d_{t+1+i} - d_{t+i}$, 有 $d_{t+1+i} = d_t + \sum_{j=0}^i \Delta d_{t+1+j}$ 所以: $E_t(d_{t+1+i}) = d_t + E_t(\sum_{j=0}^i \Delta d_{t+1+j})$

写入 $d_t - p_t$ 的后一项中:

$$\begin{aligned}
\text{后一项} &= \sum_{i=0}^{\infty} \rho^i (d_t + \sum_{j=0}^i E_t(\Delta d_{t+1+j})) \\
&= \sum_{i=0}^{\infty} \rho^i d_t + \sum_{i=0}^{\infty} \rho^i \sum_{j=0}^i E_t(\Delta d_{t+1+j}) \\
&= \frac{d_t}{1-\rho} + \sum_{j=0}^{\infty} E_t(\Delta d_{t+1+j}) \sum_{i=j}^{\infty} \rho^i \\
&= \frac{d_t}{1-\rho} + \sum_{j=0}^{\infty} \frac{\rho^j}{1-\rho} E_t(\Delta d_{t+1+j}) \\
(1-\rho) \sum_{i=0}^{\infty} \rho^i E_t d_{t+1+i} &= d_t + \sum_{j=0}^{\infty} \rho^j E_t(\Delta d_{t+1+j}) \\
d_t - p_t &= d_t - \frac{\kappa - \gamma}{1-\rho} - d_t - \sum_{j=0}^{\infty} \rho^j E_t(\Delta d_{t+1+j}) \\
&= -\frac{\kappa - \gamma}{1-\rho} - \sum_{j=0}^{\infty} \rho^j E_t(\Delta d_{t+1+j})
\end{aligned}$$

2.2 ADF test

原文中提到 “Of course, it seems likely that in practice explosive behavior in p_t may only be temporary and short-lived, as in the case of stock market bubbles that collapse after a certain period of time. Some of these possibilities can be taken into account by looking at subsamples of the data.” 直接关注整个序列的爆炸行为可能会导致错误的结果，因为爆炸行为可能只是暂时的，在一定时间后会消失。因此，可以通过分析子样本数据来考虑这些可能性，因此引入递归的 ADF 检验。

ADF 回归模型是:

$$x_t = \mu + \delta x_{t-1} + \sum_{j=1}^J \phi_j \Delta x_{t-j} + \varepsilon_t \quad (t = J + 2, \dots, n) \quad (20)$$

其中 $\Delta x_{t-j} = x_{t-j} - x_{t-j-1}$ ，是为了处理序列中的高阶自相关。

为了简化推导，或者说重点关注的是自回归项的系数，我们现在先考虑最简单的一阶 AR 模型，不加滞后差分项（即 $J = 0$ ）:

$$x_t = \mu + \delta x_{t-1} + \varepsilon_t \quad \text{for } t = 2, \dots, n \quad (21)$$

我们要对这个回归进行最小二乘估计。

设：

$$Y = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n-1} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{n-1} \end{bmatrix} \in \mathbb{R}^{(n-1) \times 2} \quad \beta = \begin{bmatrix} \mu \\ \delta \end{bmatrix} \in \mathbb{R}^2 \quad \varepsilon = \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

则回归模型为：

$$Y = X\beta + \varepsilon$$

OLS 解是：

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\mu} \\ \hat{\delta} \end{bmatrix}$$

我们只关心 $\hat{\delta}$ ，所以可以直接写：

$$\hat{\delta} = \frac{\sum_{t=2}^n (x_{t-1} - \bar{x})(x_t - \bar{x})}{\sum_{t=2}^n (x_{t-1} - \bar{x})^2}$$

这里 \bar{x} 是 x_{t-1} 的样本平均（注意如果你去掉常数项 μ ，就不用减均值）。

OLS 估计量的标准误差公式是：

$$SE(\hat{\delta}) = \sqrt{\hat{\sigma}^2 \cdot \left(\sum (x_{t-1} - \bar{x})^2 \right)^{-1}}$$

其中残差方差估计为：

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{t=2}^n (x_t - \hat{\mu} - \hat{\delta}x_{t-1})^2$$

其中 $k=2$ 是参数个数（常数项 + 自回归系数）。

当 x_t 是单位根过程（即 $\delta=1$ ，无漂移）时：

$$x_t = x_{t-1} + \varepsilon_t \quad \Rightarrow \quad x_t = \sum_{i=1}^t \varepsilon_i$$

这个过程是一个随机游走，随着样本数量 $n \rightarrow \infty$ ，我们可以将离散的过程缩放，构造连续时间极限：

$$W_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} \varepsilon_i \quad \Rightarrow \quad W(r)$$

当 $n \rightarrow \infty$, 这个过程趋近于标准布朗运动 $W(r)$, 这是 Donsker 定理的结论。
 设: 用的 OLS 回归是无截距的版本 (方便推导); 那么 ADF 统计量可以表示为:

$$ADF_n = \frac{\sum x_{t-1} \varepsilon_t}{\sqrt{\hat{\sigma}^2 \sum x_{t-1}^2}}$$

当 $\delta = 1$, 单位根情形下:

$$x_{t-1} \approx \sum_{i=1}^{t-1} \varepsilon_i \quad x_{t-1} \varepsilon_t \approx \left(\sum_{i=1}^{t-1} \varepsilon_i \right) \varepsilon_t$$

这在极限下变成了一个伊藤积分:

$$\sum_{t=1}^n x_{t-1} \varepsilon_t \rightarrow \int_0^1 W(r) dW(r)$$

而分母:

$$\sum x_{t-1}^2 \rightarrow \int_0^1 W(r)^2 dr$$

于是有极限表示:

$$ADF_n \Rightarrow \frac{\int_0^1 W(r) dW(r)}{\left(\int_0^1 W(r)^2 dr \right)^{1/2}}$$

这个极限不是标准正态分布, 而是一个非标准分布, 我们称为 Dickey-Fuller 分布。

注意: 这个结果只在单位根情形 ($\delta = 1$) 成立; 在爆炸情形 ($\delta > 1$) 下, 它有不同的极限形式, 甚至趋向于无穷大。

如果回归中带有:

截距项 μ 或者时间趋势 βt 则极限形式要对 $W(r)$ 做变换, 形成所谓的 demeaned Brownian motion 或 detrended Brownian motion:

$$\tilde{W}(r) = W(r) - \frac{1}{r} \int_0^r W(s) ds$$

这是为了去掉常数项后剩下的部分。这样写的目的, 是使得 ADF 统计量在不同模型设定下都能统一为 Brownian motion 的函数。

所以我们可以得到每个递归区间的 ADF 统计量:

$$ADF_r \Rightarrow \frac{\int_0^r \tilde{W}(s) dW(s)}{\left(\int_0^r \tilde{W}(s)^2 ds \right)^{1/2}} \quad (22)$$

我们在逐步递增样本比例的过程中, 寻找爆炸性最强的一点:

$$\sup_{r \in [r_0, 1]} ADF_r \Rightarrow \sup_{r \in [r_0, 1]} \frac{\int_0^r \tilde{W}(s) dW(s)}{\left(\int_0^r \tilde{W}(s)^2 ds \right)^{1/2}} \quad (23)$$

这个数值与标准 Dickey-Fuller t 统计量的渐近分布的右尾临界值进行对比（通过蒙特卡罗模拟得到），即可得出序列是否存在泡沫的结论。

我们还希望了解序列泡沫的开始时间和结束时间，我们就需要定义

$$\hat{r}_e = \inf_{s \geq r_0} \{s : ADF_s > cv_{\beta_n}^{\text{adf}}(s)\}, \hat{r}_f = \inf_{s \geq \hat{r}_e} \{s : ADF_s < cv_{\beta_n}^{\text{adf}}(s)\}, \quad (24)$$

因为我们使用了不断增加样本判断的方式来寻找泡沫开始和结束的区间，所以在这个式子中我们不能直接使用固定的临界值进行对比，原因在于当 $n \rightarrow \infty$ 时，犯第一类错误的概率会趋于无限大²，因此，我们需要不断严格置信水平（ β_n 不断缩小），或者使用缓速变大的 critical value，文中是定义 $cv_{\beta_n}^{\text{adf}}(s) = \log(\log(ns))/100$ 。 \hat{r}_e 定义了泡沫开始的时间，这个式子的意思是说当回归第一次出现大于临界值的情况时，定义为泡沫开始的时间。 \hat{r}_f 代表泡沫结束的时间，该式子是说当发现第一次大于临界值之后，ADF 统计量第一次小于临界值时，则为泡沫结束的时间。

2.3 Econometric analysis of explosive process

此部分利用 Phillips 和 Magdalinos（2007a, b）的工作为轻度爆炸过程提供了一种渐近分布理论，可用于当前模型的置信区间构建。

根据 Phillips and Magdalinos (2007a,b) Theorem 4.3. 对于 $y_t = \rho_n y_{t-1} + \varepsilon_t$, 其中 $\rho_n = 1 + \frac{c}{k_n}$, $c > 0$, $k_n \rightarrow \infty$, $k_n = o(n)$. 当 $n \rightarrow \infty$ 时, 有以下极限性质:

$$\begin{aligned} ((\rho_n^{-n}/k_n) \sum_{t=1}^n y_{t-1} u_t, (\rho_n^{-2n}/k_n^2) \sum_{t=1}^n y_{t-1}^2) &\Longrightarrow (XY, Y^2), \\ (k_n \rho_n^n / 2c)(\hat{\rho}_n - \rho_n) &\Longrightarrow C, \end{aligned}$$

X 和 Y 是服从 $N(0, \sigma^2/2c)$ 的独立随机变量。 $\Longrightarrow C$ 是标准柯西分布。

使用这个性质，可以直接得到第一种极限性质。

$$\frac{k_n \delta_n^n}{2c} (\hat{\delta}_n - \delta_n) \Longrightarrow C$$

在 $k_n = o(n)$ 的条件下，可以推导出第二种归一化形式

$$\frac{\delta_n^n}{\delta_n^2 - 1} (\hat{\delta}_n - \delta_n) \Longrightarrow C$$

将 $\delta_n = 1 + \frac{c}{k_n}$ 代入，那么：

$$\delta_n^2 - 1 = \left(1 + \frac{c}{k_n}\right)^2 - 1 = \frac{2c}{k_n} + \frac{c^2}{k_n^2}$$

²因为在不断增加样本进行回归的过程中，犯错的概率会随着次数的增加而提高

于是:

$$\boxed{\delta_n^2 - 1 = \frac{2c}{k_n} \left(1 + \frac{c}{2k_n}\right)} \Rightarrow \frac{1}{\delta_n^2 - 1} = \frac{k_n}{2c} \cdot \frac{1}{1 + \frac{c}{2k_n}}$$

由于 $k_n = o(n)$, 意味着 $\frac{1}{k_n} \gg \frac{1}{n}$, 所以:

$$\frac{c}{2k_n} \rightarrow 0 \quad \text{比} \frac{1}{n} \text{ 还要快} \Rightarrow \frac{1}{1 + \frac{c}{2k_n}} \rightarrow 1$$

因此可以直接近似:

$$\frac{1}{\delta_n^2 - 1} = \frac{k_n}{2c} \cdot (1 + o(1))$$

我们从第一种归一化形式开始:

$$\frac{k_n \delta_n^n}{2c} (\hat{\delta}_n - \delta_n)$$

我们用上面公式中的近似替换 $\frac{k_n}{2c}$, 就得到:

$$= \delta_n^n \cdot \frac{1}{\delta_n^2 - 1} \cdot (1 + o(1)) \cdot (\hat{\delta}_n - \delta_n)$$

合并得到:

$$= \left(\frac{\delta_n^n}{\delta_n^2 - 1} (\hat{\delta}_n - \delta_n) \right) (1 + o(1))$$

而因为 $(1 + o(1)) \rightarrow 1$, 不会影响极限分布 (Cauchy 分布是闭集稳定分布), 所以仍有:

$$\frac{\delta_n^n}{\delta_n^2 - 1} (\hat{\delta}_n - \delta_n) \Longrightarrow C$$

经过相应的移项处理, 可以得到:

$$\left(\hat{\delta}_n \pm \frac{(\hat{\delta}_n)^2 - 1}{(\hat{\delta}_n)^n} C_\alpha \right), \quad (25)$$

我们就可以使用标准柯西分布的临界值构造置信区间。

2.4 Finite sample bias correction via indirect inference

问题的产生:

$$y_t = \delta_n x_t + \mu_t \quad (26)$$

OLS 得到 $\hat{\delta}$

$$\hat{\delta} = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2} = \frac{\sum y_{t-1}(\delta y_{t-1} + \mu_t)}{\sum y_{t-1}^2} = \delta + \frac{\sum y_{t-1} \mu_t}{\sum y_{t-1}^2}$$

求期望, 许多文献认为这个偏误是向下偏误的:

$$E(\hat{\delta}) = \delta + \underbrace{E\left(\frac{\sum_{t=1}^n E y_{t-1} \mu_t}{\sum_{t=1}^n y_{t-1}^2}\right)}_{\text{偏差来源}}$$

在时间序列分析中, 用最小二乘法 (Least Squares, LS) 估计 δ 时, 会得到一个估计值 δ^{LS} 。但当样本量 n 不够大时, 这个估计值是有偏的 (biased), 也就是说 δ^{LS} 的平均值不等于真实的 δ_0 。

传统方法: 有些方法, 比如 Kendall (1954) 提出的程序, 试图通过一个明确的公式来校正这种偏误。但这种方法需要事先知道偏误的具体形式。

间接推断的优势: 间接推断 (Indirect Inference) 不需要明确的偏误公式, 而是通过模拟来估计和校正偏误。这是一种更灵活的方法, 尤其适用于复杂的模型。间接推断的核心是通过模拟数据来“校准”偏误, 找到更准确的 δ 估计值。以下是具体步骤:

步骤 1: 对观测数据应用最小二乘法

用真实观测数据 $\{x_t; t \leq n\}$ 计算最小二乘估计 δ^{LS} 。这个估计是有偏的。

步骤 2: 模拟时间序列路径假设一个 δ 值 (可以是任何候选值)。

根据模型 $x_t = \mu_x + \delta x_{t-1} + \varepsilon_{xt}$ 和误差分布 $\varepsilon_{xt} \sim \text{NID}(0, \sigma_x^2)$, 生成 H 条模拟的时间序列路径, 每条路径记为 $\{x_t^h(\delta); t \leq n\}$ (其中 $h = 1, \dots, H$)。

对每条模拟路径, 用最小二乘法计算一个估计值 $\delta_h(\delta)$ 。

步骤 3: 定义间接推断估计量

间接推断估计量 $\delta_{n,H}^{II}$ 是通过以下方式定义的:

$$\delta_{n,H}^{II} = \arg \min_{\delta \in \Phi} \left\| \delta^{LS} - \frac{1}{H} \sum_{h=1}^H \delta_h(\delta) \right\|,$$

$\|\cdot\|$ 是一个距离度量 (比如欧几里得距离),

Φ 是参数空间 (一个有限的范围),

$\frac{1}{H} \sum_{h=1}^H \delta_h(\delta)$ 是 H 个模拟估计的平均值。

意思是: 找到一个 δ , 使得观测到的 δ^{LS} 和模拟估计的平均值尽可能接近。

当模拟次数 H 非常大 (趋于无穷) 时, 平均值 $\frac{1}{H} \sum_{h=1}^H \delta_h(\delta)$ 会趋近于期望值 $q_n(\delta) = E(\delta_h(\delta))$, 这个期望值被称为“绑定函数” (binding function)。这时, 间接推断估计量变为:

$$\delta_n^{II} = \arg \min_{\delta \in \Phi} \left\| \delta^{LS} - q_n(\delta) \right\|.$$

如果 $\|\cdot\|$ 是一个标准的距离度量 (比如欧几里得距离), 那么“距离最小”通常意味着:

$$\delta^{LS} - q_n(\delta) = 0,$$

因为当两者相等时，距离为 0，这是最小的可能值。因此，最优的 δ 满足：

$$q_n(\delta) = \delta^{LS}.$$

当 $q_n(\delta)$ 可逆时，有：

$$\delta = q_n^{-1}(\delta^{LS}).$$

这个 δ 就是使距离最小的值，也就是 δ_n^{II} 。

也就是说如果 $q_n(\delta)$ 是可逆的（即可以求反函数），我们可以直接计算：

$$\delta_n^{II} = q_n^{-1}(\delta_n^{LS}),$$

换言之，通过将 δ^{LS} 代入绑定函数的反函数，直接得到间接推断估计。

2.5 Estimating origination and collapse dates

$$\lim_{n \rightarrow \infty} P(ADF_s > cv_{\beta_n}^{\text{adf}}(s)) = P\left(\frac{\int_0^s \widetilde{W} dW}{\left(\int_0^s \widetilde{W}^2\right)^{1/2}} = \infty\right) = 0. \quad (27)$$

观察这个式子，我们可以知道，当样本数量趋于无穷时，拒绝单位根原假设判断为泡沫根的概率趋于 0，为了估计出泡沫起始和结束时间，我们必须指定一个模型，允许在单位根和轻度爆炸事件之间切换。

$$\begin{aligned} x_t &= x_{t-1} 1\{t < \tau_e\} + \delta_n x_{t-1} 1\{\tau_e \leq t \leq \tau_f\} \\ &+ \left(\sum_{k=\tau_f+1}^t \varepsilon_k + x_{\tau_f}^* \right) 1\{t > \tau_f\} + \varepsilon_t 1\{t \leq \tau_f\} \\ \delta_n &= 1 + \frac{c}{n^\alpha}, c > 0, \alpha \in (0, 1), \end{aligned} \quad (28)$$

这是一个分段时间序列模型，描述了数据在不同阶段的行为：

$t < \tau_e$ ：在爆炸性行为开始之前，模型是单位根过程： $x_t = x_{t-1} + \varepsilon_t$ 。这表示数据是非平稳的，但没有爆炸性增长。

$\tau_e \leq t \leq \tau_f$ ：在爆炸性行为期间（从 τ_e 到 τ_f ），模型变为 $x_t = \delta_n x_{t-1} + \varepsilon_t$ ，其中 $\delta_n > 1$ ，表示爆炸性增长（因为系数大于 1）。

$t > \tau_f$ ：爆炸性行为结束后，模型变为 $x_t = x_{\tau_f}^* + \sum_{k=\tau_f+1}^t \varepsilon_k$ ，表示从爆炸性结束时的值 $x_{\tau_f}^*$ 开始，数据重新变成一个随机游走（累积噪声）。

噪声项: $\varepsilon_t 1\{t \leq \tau_f\}$ 表示噪声只在爆炸性行为结束前直接影响模型。

δ_n 的定义: $\delta_n = 1 + \frac{c}{n^\alpha}$, 其中 $c > 0$, $\alpha \in (0, 1)$ 。

这表示爆炸性行为的强度是“局部”的 (locally explosive), 因为 $\delta_n \rightarrow 1$ (当 $n \rightarrow \infty$ 时), 但在有限样本中 $\delta_n > 1$, 足以产生爆炸性行为。

参数 c 控制爆炸性的强度, α 控制 δ_n 接近 1 的速度。

这个模型刻画了一次爆炸性行为 (single explosive episode), 即从 τ_e 到 τ_f , 允许数据从单位根 (平稳增长) 切换到爆炸性行为, 再回到非爆炸性状态。

这样我们就可以在一个子样本中寻找泡沫根:

$$\hat{r}_e = \inf_{s \geq r_0} \{s : ADF_s > cv_{\beta_n}^{\text{adf}}(s)\}, \quad (29)$$

下面需要讨论在每个子样本中检验的势如何:

考虑如下的模型:

$$x_t = \delta_n x_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} (0, \sigma^2),$$

其中 $\delta_n = 1 + \frac{c}{n^\alpha}$, $0 < \alpha < 1$, 表示轻微爆炸过程。

我们只使用样本区间 $t = \tau_e, \dots, \tau$ 来估计 δ_n , 记样本大小为 $T = \tau - \tau_e + 1 \sim O(n)$ 。

从定义出发, 区间 $[\tau_e, \tau]$ 上对 δ_n 的 OLS 估计为:

$$\hat{\delta}_n(\tau) = \frac{\sum_{t=\tau_e}^{\tau} x_{t-1} x_t}{\sum_{t=\tau_e}^{\tau} x_{t-1}^2} = \frac{\sum_{t=\tau_e}^{\tau} x_{t-1} (\delta_n x_{t-1} + \varepsilon_t)}{\sum_{t=\tau_e}^{\tau} x_{t-1}^2} = \delta_n + \frac{\sum_{t=\tau_e}^{\tau} x_{t-1} \varepsilon_t}{\sum_{t=\tau_e}^{\tau} x_{t-1}^2}.$$

因此, 估计误差项为:

$$\hat{\delta}_n(\tau) - \delta_n = \frac{\sum_{t=\tau_e}^{\tau} x_{t-1} \varepsilon_t}{\sum_{t=\tau_e}^{\tau} x_{t-1}^2}.$$

考虑如下归一化误差项:

$$\frac{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}{2c} (\hat{\delta}_n(\tau) - \delta_n).$$

这个标准化因子的选择来源于 Phillips and Magdalinos(2007) Theorem 4.3:

在轻微爆炸过程中:

$$\frac{k_n \rho_n^n}{2c} (\hat{\rho}_n - \rho_n) \Rightarrow C, \quad \text{其中 } k_n = n^{(1+\alpha)/2}.$$

我们这里考虑的是局部样本 $\tau_e \sim \tau$, 而不是整个样本 $1 \sim n$, 所以把 ρ_n^n 替换成 $\delta_n^{\tau-\tau_e}$, 从而构造出适应局部样本的标准化因子。

即我们定义:

$$k_T = n^{(1+\alpha)/2}, \quad \text{且} \quad \rho_n^n \rightarrow \delta_n^{\tau-\tau_e}.$$

所以我们尝试研究的是局部样本下:

$$\frac{k_T \delta_n^{\tau-\tau_e}}{2c} (\hat{\delta}_n(\tau) - \delta_n).$$

回到我们刚刚推导的 OLS 误差项:

$$\hat{\delta}_n - \delta_n = \frac{\sum x_{t-1} \varepsilon_t}{\sum x_{t-1}^2}.$$

带入标准化项:

$$\frac{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}{2c} (\hat{\delta}_n - \delta_n) = \frac{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}{2c} \cdot \frac{\sum_{j=\tau_e}^{\tau} x_{j-1} \varepsilon_j}{\sum_{j=\tau_e}^{\tau} x_{j-1}^2}.$$

我们希望把分子分母都标准化, 以便分别研究其极限。

观察:

分子为: $\sum x_{j-1} \varepsilon_j$ 分母为: $\sum x_{j-1}^2$

考虑对分子除以:

$$n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}, \quad \text{得到} \quad \frac{\sum x_{j-1} \varepsilon_j}{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}.$$

而分母我们除以:

$$n^{1+\alpha} \delta_n^{2(\tau-\tau_e)}, \quad \text{得到} \quad \frac{\sum x_{j-1}^2}{n^{1+\alpha} \delta_n^{2(\tau-\tau_e)}}.$$

于是可以写出:

$$\frac{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}{2c} (\hat{\delta}_n - \delta_n) = \frac{\frac{\sum x_{j-1} \varepsilon_j}{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}}{\frac{2c \sum x_{j-1}^2}{n^{1+\alpha} \delta_n^{2(\tau-\tau_e)}}}.$$

再将其改写, 得到:

$$= \frac{\frac{\delta_n^{-(\tau-\tau_e)}}{n^{(1+\alpha)/2}} \sum x_{j-1} \varepsilon_j \cdot \{1 + o_p(1)\}}{\frac{2c \delta_n^{-2(\tau-\tau_e)}}{n^{1+\alpha}} \sum x_{j-1}^2 \cdot \{1 + o_p(1)\}}.$$

我们现在应用 Phillips and Magdalinos(2007) Theorem 4.3(a):

$$\left(\frac{\rho_n^{-n}}{k_n} \sum y_{t-1} u_t, \frac{\rho_n^{-2n}}{k_n^2} \sum y_{t-1}^2 \right) \Rightarrow (XY, Y^2),$$

局部版本为:

$$\left(\frac{\delta_n^{-(\tau-\tau_e)}}{n^{(1+\alpha)/2}} \sum x_{j-1} \varepsilon_j, \frac{\delta_n^{-2(\tau-\tau_e)}}{n^{1+\alpha}} \sum x_{j-1}^2 \right) \Rightarrow (XY, Y^2).$$

那么他们的比值收敛到:

$$\frac{XY}{2cY^2} = \frac{X}{2cY} = C.$$

因此:

$$\frac{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}{2c} (\hat{\delta}_n - \delta_n) \Rightarrow C.$$

这个式子是分析极限过程中的 OLS 估计量与单位根的偏差在 mild explosiveness 下的行为, 具体来说, 是对:

$$\tau(\hat{\delta}_n(\tau) - 1)$$

这个项在 $n \rightarrow \infty$ 下的极限进行展开和评估。我们一行一行分析这个式子。

第一行: 分解估计误差

$$\tau(\hat{\delta}_n(\tau) - 1) = \tau(\hat{\delta}_n(\tau) - \delta_n) + \tau(\delta_n - 1)$$

这是一个基本恒等式, 把估计误差拆成了两部分:

1. 估计量和真值之间的误差 ($\hat{\delta}_n(\tau) - \delta_n$);
2. 真值和单位根之间的偏差 ($\delta_n - 1$);

第二行: 近似代入渐近阶数

$$\tau(\delta_n - 1) + o_p\left(\frac{\tau}{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}\right)$$

其中:

$\tau(\delta_n - 1)$ 是主导项;

$\hat{\delta}_n(\tau) - \delta_n$ 这项之前我们讨论过, 它是:

$$\hat{\delta}_n(\tau) - \delta_n = O_p\left(\frac{1}{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}\right)$$

乘上 τ 后成为:

$$o_p\left(\frac{\tau}{n^{(1+\alpha)/2} \delta_n^{\tau-\tau_e}}\right)$$

第三行: 用 Mild Explosive 假设代入表达式

根据 mild explosiveness 的假设:

$$\delta_n = 1 + \frac{c}{n^\alpha}, \quad \tau = [nr] \sim nr,$$

所以:

$$\tau(\delta_n - 1) = nr \cdot \frac{c}{n^\alpha} = n^{1-\alpha}rc.$$

因此我们得到:

$$\tau(\hat{\delta}_n(\tau) - 1) = n^{1-\alpha}rc + o_p(1).$$

最后一项: 结论是发散的

因为 $\alpha \in (0, 1)$, 所以 $1 - \alpha > 0$, 因此:

$$n^{1-\alpha}rc \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

所以最终结果是:

$$\tau(\hat{\delta}_n(\tau) - 1) \rightarrow \infty \quad \text{in probability.}$$

它表明, 在 mildly explosive 模型下, 即使 $\delta_n \rightarrow 1$, 我们使用 OLS 估计得到的 $\hat{\delta}_n$ 在有限时间点 τ 下, 仍然导致:

$$\tau(\hat{\delta}_n(\tau) - 1) \rightarrow \infty$$

也就是说, 估计量相对于单位根存在显著偏离, 即便爆炸程度很弱。这个性质使得 unit root 检验可以在 mildly explosive 过程中区分 $\delta_n > 1$ 的情形。

我们还需要就这个子样本的估计值计算出相应的 t 统计量。下面的式子本质上是在构造一个 t 统计量, 用以检验 $\delta_n = 1$ (即单位根假设) 在轻微爆炸 (mildly explosive) 环境下的偏离速度。

一、式子原型与含义

我们关心的统计量是

$$\left(\frac{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} (\hat{\delta}_n(\tau) - 1). \quad (30)$$

它等价于

$$\frac{\left(\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2 \right)^{1/2}}{\hat{\sigma}_\tau} \cdot (\hat{\delta}_n(\tau) - 1). \quad (31)$$

其中, $\hat{\delta}_n(\tau)$ 是在样本区间 $1, \dots, \tau$ (或更一般地, 从爆炸开始点 τ_e 到 τ) 用 OLS 拟合回归

$$x_t = \delta x_{t-1} + \varepsilon_t \quad (32)$$

得到的估计值。检验 $\delta = 1$ (单位根) 时, 常用的做法是构造

$$\frac{\hat{\delta}_n(\tau) - 1}{\text{估计的标准误}} \quad (33)$$

如果把 OLS 回归的残差方差估计记为 $\hat{\sigma}_\tau^2$, 则 OLS “标准误” 大致与 $\hat{\sigma}_\tau / (\sum x_{j-1}^2)^{1/2}$ 同阶。因此

$$\frac{\hat{\delta}_n(\tau) - 1}{\hat{\sigma}_\tau / \sqrt{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}} = \sqrt{\frac{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2}} (\hat{\delta}_n(\tau) - 1) \quad (34)$$

就是一个“t 统计量”形式。我们要研究这个 t 统计量在 $\delta_n = 1 + c/n^\alpha$ 这类轻微爆炸过程中, 如何发散到无穷, 以及它的发散速度。

二、整体分两步拆解

1. 先把 $\hat{\delta}_n(\tau) - 1$ 展开

已知 (见前面推导或 Phillips-Magdalinos 框架):

$$\tau (\hat{\delta}_n(\tau) - 1) = n^{1-\alpha} r c + o_p(1), \quad \text{其中 } \tau = \lfloor nr \rfloor, r \in (0, 1). \quad (35)$$

也就是说, $\hat{\delta}_n(\tau) - 1 = O_p(n^{-\alpha})$, 但乘以 τ 后得到发散速度:

$$\tau (\hat{\delta}_n(\tau) - 1) \approx n^{1-\alpha} r c. \quad (36)$$

2. 再把“设计矩阵范数”与“残差方差”的估算插入

\tilde{x}_{j-1} 通常指回归时自变量的中心化或去趋势残差, 但在轻微爆炸文献里常直接用 x_{t-1} 。在爆炸过程里, \tilde{x}_{j-1} 的数量级由 x_{τ_e} (爆炸开始时刻的值) 和“爆炸倍数” $\delta_n^{j-1-\tau_e}$ 决定。

$\hat{\sigma}_\tau^2$ 是 ε_t 的残差方差估计。因为 ε_t 是均值 0、方差 σ^2 的噪声, 在轻微爆炸区间里, 残差项数量级不会被 δ_n 放大太多, 所以 $\hat{\sigma}_\tau^2$ 会收敛到某个常数倍, 具体推导见下。

三、关键近似: $\sum \tilde{x}_{j-1}^2$ 和 $\hat{\sigma}_\tau^2$

下面给出两项的渐近近似 (所有 $o_p(1)$ 都在最后统一标为 $\{1 + o_p(1)\}$):

1. 设计矩阵二范数 $\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2$ 的近似

在轻微爆炸框架下, 从爆炸起点 τ_e 开始, x_{j-1} 会被 $\delta_n^{j-1-\tau_e}$ 放大。近似为

$$\tilde{x}_{j-1} = \delta_n^{j-1-\tau_e} \tilde{x}_{\tau_e} \quad (37)$$

于是

$$\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2 = \sum_{j=\tau_e}^{\tau} [\delta_n^{j-1-\tau_e} \tilde{x}_{\tau_e}]^2 = \tilde{x}_{\tau_e}^2 \sum_{k=0}^{\tau-\tau_e} \delta_n^{2k} \quad (38)$$

当 $\delta_n = 1 + \frac{c}{n^\alpha}$ 且 $\tau - \tau_e = O(n)$ 时，几何级数近似为

$$\sum_{k=0}^{\tau-\tau_e} \delta_n^{2k} = \frac{\delta_n^{2(\tau-\tau_e)+2} - 1}{\delta_n^2 - 1} \approx \frac{\delta_n^{2(\tau-\tau_e)}}{2(\delta_n - 1)} \quad (39)$$

而 $\delta_n - 1 = cn^{-\alpha}$ ，因此

$$\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2 = \frac{n^\alpha \delta_n^{2(\tau-\tau_e)}}{2c} \tilde{x}_{\tau_e}^2 \{1 + o_p(1)\} \quad (40)$$

2. 残差方差估计 $\hat{\sigma}_\tau^2$ 的近似

OLS 残差方差估计

$$\hat{\sigma}_\tau^2 = \frac{1}{\tau - 1} \sum_{j=1}^{\tau} (x_j - \hat{\delta}_n(\tau) x_{j-1})^2 \approx \frac{1}{\tau} \sum_{j=1}^{\tau} \varepsilon_j^2 \quad (41)$$

在轻微爆炸框架下，考虑爆炸部分的影响，等价结果为

$$\hat{\sigma}_\tau^2 = \frac{\delta_n^{2(\tau-\tau_e)}}{\tau^2} \frac{r_e}{c^2 r} \tilde{x}_{\tau_e}^2 \{1 + o_p(1)\} \quad (42)$$

其中 $r_e = \tau_e/n$ 。

四、把两者代入 t 统计量

1. 先写成“乘以 τ ”的形式

原式可写为

$$\left(\frac{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} (\hat{\delta}_n(\tau) - 1) = \left(\frac{\tau^{-2} \sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} \tau (\hat{\delta}_n(\tau) - 1) \quad (43)$$

其中

$$\tau (\hat{\delta}_n(\tau) - 1) = n^{1-\alpha} r c \{1 + o_p(1)\} \quad (44)$$

2. 将 $\sum \tilde{x}^2$ 与 $\hat{\sigma}_\tau^2$ 的渐近形式带入

分子部分（除以 τ^2 后）：

$$\frac{\tau^{-2} \sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} = \frac{n^\alpha c r}{2 r_e} \{1 + o_p(1)\} \quad (45)$$

开平方得

$$\left(\frac{\tau^{-2} \sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} = n^{\alpha/2} \frac{\sqrt{c r}}{\sqrt{2 r_e}} \{1 + o_p(1)\} \quad (46)$$

3. 和 $\tau(\hat{\delta}_n(\tau) - 1)$ 相乘

$$\begin{aligned}
& \left(\frac{\tau^{-2} \sum \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} \times \tau(\hat{\delta}_n(\tau) - 1) \\
&= n^{\alpha/2} \frac{\sqrt{cr}}{\sqrt{2r_e}} \{1 + o_p(1)\} \times n^{1-\alpha} r c \{1 + o_p(1)\} \\
&= n^{1-\alpha/2} \frac{c^{3/2} r^{3/2}}{\sqrt{2r_e}} \{1 + o_p(1)\}
\end{aligned} \tag{47}$$

最终得到

$$\boxed{\left(\frac{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} (\hat{\delta}_n(\tau) - 1) = n^{1-\alpha/2} \frac{c^{3/2} r^{3/2}}{\sqrt{2r_e}} \{1 + o_p(1)\}} \tag{48}$$

五、整个式子的含义

左侧

$$\left(\frac{\sum_{j=1}^{\tau} \tilde{x}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} (\hat{\delta}_n(\tau) - 1) \tag{49}$$

就是一个“t 统计量”形式：分母是残差标准差，分子是 $\hat{\delta}_n(\tau) - 1$ 乘以设计矩阵范数，再开平方以匹配 OLS 标准误的量纲。实际意义是：在时刻 τ 前用 OLS 估计出的 $\hat{\delta}$ 和单位根 1 之间的差异，经由这个标准误放大（或缩放）之后，会得到一个统计量。

右侧

$$n^{1-\alpha/2} \frac{c^{3/2} r^{3/2}}{\sqrt{2r_e}} \{1 + o_p(1)\} \tag{50}$$

则告诉我们：

1. 当 $n \rightarrow \infty$ 时，只要 $0 < \alpha < 1$ ，就有 $1 - \alpha/2 > 0$ ，因此这整个 t 统计量会以阶 $n^{1-\alpha/2}$ 的速度发散到正无穷。

2. 前面的常数 $\frac{c^{3/2} r^{3/2}}{\sqrt{2r_e}}$ 反映了：

c 是轻微爆炸系数 ($\delta_n = 1 + c/n^\alpha$),

$r = \lim \tau/n$ 是总样本占比,

$r_e = \lim \tau_e/n$ 是爆炸开始点占比。

换句话说：在 mild explosiveness 设置下，即使 δ_n 只比 1 略大一丁点 ($\delta_n - 1 = cn^{-\alpha}$)，我们构造的这个“学生化残差统计量”也会按 $n^{1-\alpha/2}$ 的速率发散。这说明：在这种情况下 OLS 回归几乎“肯定”会拒绝 $\delta = 1$ 的单位根假设，并且这个拒绝的力度（“t 值”的大小）与样本量 n 、爆炸程度 α 、 c 等参数有关。

(19) 式：对于所有 $\tau = [nr]$ 且 $r > r_e$ ，有

$$P(\text{ADF}_\tau > c_{\beta_n}^{\text{adf}}(r)) \rightarrow 1, \quad \text{条件是 } c_{\beta_n}^{\text{adf}}(r) n^{1-\alpha/2} \rightarrow 0.$$

DF 统计量大致上

$$\text{ADF}_\tau \approx O_p(n^{1-\alpha/2}).$$

换句话说，当 $r > r_e$ （真正处在爆炸期）时， ADF_τ 的 ** 数量级为 $n^{1-\alpha/2}$ ，并且实际上 $\text{ADF}_\tau/n^{1-\alpha/2}$ 收敛到一个正的常数（不是退化到 0，也不是发散到 ∞ ）。因此，只要我们选择的批判值

$$c_{\beta_n}^{\text{adf}}(r) = o(n^{1-\alpha/2}), \quad \text{即 } c_{\beta_n}^{\text{adf}}(r) n^{1-\alpha/2} \rightarrow 0,$$

那么

$$\frac{\text{ADF}_\tau}{c_{\beta_n}^{\text{adf}}(r)} = \frac{O_p(n^{1-\alpha/2})}{o(n^{1-\alpha/2})} \xrightarrow{p} +\infty,$$

从而

$$P(\text{ADF}_\tau > c_{\beta_n}^{\text{adf}}(r)) = P\left(\frac{\text{ADF}_\tau}{c_{\beta_n}^{\text{adf}}(r)} > 1\right) \rightarrow 1.$$

这就“直观地”给出了 (19) 式的推导逻辑：

当 $r > r_e$ 时， ADF_τ 是 Order $n^{1-\alpha/2}$ 的大数；只要批判值的增长“比 $n^{1-\alpha/2}$ 慢”（即 $c_{\beta_n}^{\text{adf}}(r) n^{1-\alpha/2} \rightarrow 0$ ），就必然出现 ADF_τ 大大超过 $c_{\beta_n}^{\text{adf}}(r)$ 的情形，概率趋向 1。

真实的“爆炸起始点”是 r_e ，对应整数 $\tau_e = \lfloor nr_e \rfloor$ 。我们想要证明：

$$\hat{r}_e \xrightarrow{p} r_e, \quad n \rightarrow \infty,$$

即“估计的爆炸起始点”概率收敛到真正的 r_e 。

要说明这一点，通常分两步做：

1. 不可能过早（under-estimate）——证明 $P(\hat{r}_e < r_e) \rightarrow 0$ 。
2. 不可能太晚（over-estimate）——证明 $P(\hat{r}_e > r_e + \eta) \rightarrow 0$ ，对任意微小 $\eta > 0$ 。

证明 $P(\hat{r}_e < r_e) \rightarrow 0$

我们先看在“ $\tau = \lfloor nr \rfloor$ ”，若 $r < r_e$ ，此时 x_t 过程还接近单位根（或弱近似单位根），ADF 统计量在这种区间内不会像爆炸区间那样快速发散。具体来讲：

1. 当 $r < r_e$ 时，对应的整数时刻 $\tau < \tau_e$ ，实际上这部分样本属于“局部平稳 / 单位根区间”。已知在单位根或近单位根区间下（如果 $\delta_n = 1 + O(n^{-\alpha})$ 且样本长度还在“非爆炸期”）， ADF_τ 有一个弱极限，例如趋向某个随机变量分布（不发散到 ∞ ）。

2. 同时，我们要求 critical value 序列 $c_{\beta_n}^{\text{adf}}(r)$ “必须发散到 ∞ ”（以确保 asymptotic type-I error 可忽略）。也就是说，对于每一个固定的 $r < r_e$ ，

$$c_{\beta_n}^{\text{adf}}(r) \rightarrow \infty, \quad \text{而且} \quad c_{\beta_n}^{\text{adf}}(r) n^{1-\alpha/2} \rightarrow 0.$$

这里“趋于 ∞ ”保证“在纯单位根 / 近单位根期，ADF 统计量（概率弱收敛到某分布）”不会轻易超过“越来越大的临界值”。

3. 因此，对任意 $r < r_e$ ，随着 $n \rightarrow \infty$ ，

$$P(\text{ADF}_\tau > c_{\beta_n}^{\text{adf}}(r)) \rightarrow 0.$$

这意味着在时刻 $r < r_e$ 区间几乎不可能出现“ADF-越界”，从而

$$P(\hat{r}_e < r_e) = P(\exists s < r_e : \text{ADF}_s > c_{\beta_n}^{\text{adf}}(s)) \leq \sum_{s < r_e} P(\text{ADF}_s > c_{\beta_n}^{\text{adf}}(s)) \rightarrow 0.$$

综上所述可得：

$$P(\hat{r}_e < r_e) \rightarrow 0.$$

证明 $P(\hat{r}_e > r_e + \eta) \rightarrow 0$

接下来我们要证明：对任意固定的小常数 $\eta > 0$ ，都有

$$P(\hat{r}_e > r_e + \eta) \rightarrow 0.$$

换句话说：不可能“检测爆炸”永远拖到比真实 r_e 晚至少 η 才出现。证明思路就是利用 (19) 式：一旦真正进入爆炸区间——也就是任何 $r > r_e$ ——ADF 统计量以 $n^{1-\alpha/2}$ 速率发散，超过临界值的概率 $\rightarrow 1$ 。

1. 给定任意 $\eta > 0$ ，取 $r' = r_e + \eta$ 。考虑对应的整数时刻 $\tau' = \lfloor nr' \rfloor$ 。

2. 只要批判值满足 (19) 所要求的条件

$$c_{\beta_n}^{\text{adf}}(r') n^{1-\alpha/2} \rightarrow 0,$$

就可以从 (19) 得到

$$P(\text{ADF}_{\tau'} > c_{\beta_n}^{\text{adf}}(r')) \rightarrow 1.$$

3. 如果 $\hat{r}_e > r_e + \eta$ ，意味着“在所有时刻 $s \leq r_e + \eta$ （向下取整到整数索引以后）， ADF_s 都没有超过对应的临界值”，特别地，对应整数 τ' 时刻， $\text{ADF}_{\tau'} \leq c_{\beta_n}^{\text{adf}}(r')$ 。所以

$$\{\hat{r}_e > r_e + \eta\} \subset \{\text{ADF}_{\tau'} \leq c_{\beta_n}^{\text{adf}}(r')\}.$$

于是

$$P(\hat{r}_e > r_e + \eta) \leq P(\text{ADF}_{\tau'} \leq c_{\beta_n}^{\text{adf}}(r')) = 1 - P(\text{ADF}_{\tau'} > c_{\beta_n}^{\text{adf}}(r')) \rightarrow 0.$$

这完成了第二个方向的证明。

因此

$$P(\hat{r}_e > r_e + \eta) \rightarrow 0 \quad \forall \eta > 0.$$

推出 $\hat{r}_e \xrightarrow{p} r_e$

结合两条结果，就可得出：

$$P(|\hat{r}_e - r_e| > \eta) = P(\hat{r}_e < r_e - \eta) + P(\hat{r}_e > r_e + \eta) \rightarrow 0 + 0 = 0,$$

对任意 $\eta > 0$ 成立。这正是所谓 $\hat{r}_e \xrightarrow{p} r_e$ 。

这里有一个必要前提，就是：

$$\underbrace{\frac{1}{c_{\beta_n}^{\text{adf}}(r)}}_{\rightarrow 0} + \underbrace{\frac{c_{\beta_n}^{\text{adf}}(r)}{n^{1-\alpha/2}}}_{\rightarrow 0} \rightarrow 0 \quad \text{对所有 } r \in [r_0, 1].$$

这条 (20) 的含义就是：

1. $\frac{1}{c_{\beta_n}^{\text{adf}}(r)} \rightarrow 0$ ：批判值必须发散到 ∞ ，以保证在“非爆炸期” ($r < r_e$) 里， ADF_τ 不会误判越界。
2. $\frac{c_{\beta_n}^{\text{adf}}(r)}{n^{1-\alpha/2}} \rightarrow 0$ ：批判值相对 ADF_τ （在爆炸期）的发散速度要慢，否则当真正进入 $r > r_e$ 时，若批判值反而比 ADF_τ 生长得更快，就可能在爆炸期仍旧无法越界，导致“延迟”检出爆炸。

综上，(20) 保证了两边（“过早”和“过晚”）都不会出错，从而递归 ADF 的“首次显著”时点 \hat{r}_e 能以概率收敛的方式锁定真实的爆炸起始点 r_e 。

(20) 式的推导与含义

(20) 式：对所有 $r \in [r_0, 1]$ ，要有

$$\frac{1}{c_{\beta_n}^{\text{adf}}(r)} + \frac{c_{\beta_n}^{\text{adf}}(r)}{n^{1-\alpha/2}} \rightarrow 0.$$

这条条件本质上就是对批判值 $c_{\beta_n}^{\text{adf}}(r)$ 的 ** 生长速度 ** 提出了“夹在两端”之间的要求：

1. 必须趋于 ∞ （即 $\frac{1}{c_{\beta_n}^{\text{adf}}(r)} \rightarrow 0$ ），以便保证在“ $r < r_e$ ”（单位根近似区间）时，ADF 统计量不会因为批判值过小而误拒；

2. 又不能增长得太快，必须比 $n^{1-\alpha/2}$ 慢（即 $\frac{c_{\beta_n}^{\text{adf}}(r)}{n^{1-\alpha/2}} \rightarrow 0$ ），这样一旦真正进入“ $r > r_e$ ”爆炸区间，ADF 统计量 $O_p(n^{1-\alpha/2})$ 才能“赶得上并超过”批判值，避免“延后”检出。

换一种说法，(20) 要求：

$$c_{\beta_n}^{\text{adf}}(r) \rightarrow \infty, \quad c_{\beta_n}^{\text{adf}}(r) = o(n^{1-\alpha/2}).$$

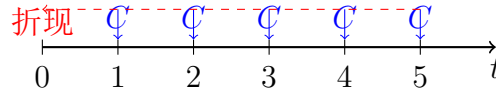
一个常见的满足该要求的例子是

$$c_{\beta_n}^{\text{adf}}(r) = c \ln \ln(n), \quad 0 < c < \infty, \quad \text{对所有 } r \in [r_0, 1],$$

因为 $\ln \ln(n) \rightarrow \infty$ 但 $\ln \ln(n)/n^{1-\alpha/2} \rightarrow 0$ 。事实上，文中提到“任何缓变（slowly varying）的增长，例如 $c(\ln \ln(n))^2$ 等”都可以，只要满足 $\ln \ln(n) = o(n^{1-\alpha/2})$ 。

3 数学基础

3.1 现金流折现以及现值计算

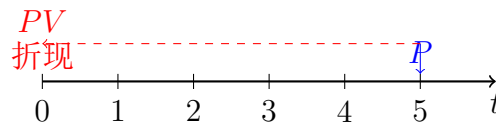


现金流折现过程：

每期末支付 C ，共 n 期，利率 r 。每期现金流 C 在 $t = k$ 折现到 $t = 0$ ，折现值为 $\frac{C}{(1+r)^k}$ 。总现值：

$$PV = C \cdot \frac{1 - (1+r)^{-n}}{r}.$$

解释：时间线显示现金流在 $t = 1, 2, \dots, n$ 处发生，通过折现因子 $(1+r)^{-k}$ ，将其折现到 $t = 0$ ，红色虚线表示折现方向。



现值计算：

未来价格 P 在 $t = n$ 发生，利率 r 。折现公式：

$$PV = P \cdot (1+r)^{-n}.$$

解释：时间线显示未来价格 P 在 $t = n$ 处发生，通过折现因子 $(1+r)^{-n}$ ，将其折现到 $t = 0$ ，红色虚线表示折现方向，最终在 $t = 0$ 得到现值 PV 。

3.2 范数

$\|\cdot\|$ 通常表示一个范数 (norm)，在数学和统计学中，它衡量向量或点之间的距离。常见的范数包括：

欧几里得范数 (Euclidean norm)： $\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ ，这是二维或更高维空间中最直观的距离。

曼哈顿范数 (Manhattan norm)： $\|x\| = |x_1| + |x_2| + \cdots + |x_n|$ ，表示沿坐标轴移动的总距离。

最大范数 (Maximum norm)： $\|x\| = \max(|x_1|, |x_2|, \dots, |x_n|)$ ，取各分量绝对值的最大值。

3.3 o_p 和 O_p 符号

在概率论和统计学中，特别是在渐进分析 (asymptotic analysis) 中， o_p 和 O_p 是描述随机变量序列渐进行为的重要符号。它们是经典的“小 o ”和“大 O ”符号的概率版本，用于表示随机变量的相对大小或增长速率。

1. 定义

O_p (大 O 概率符号) 一个随机变量序列 X_n 满足 $X_n = O_p(a_n)$ ，如果对于任意 $\epsilon > 0$ ，存在常数 $M < \infty$ ，使得当 n 足够大时：

$$P\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon.$$

直观理解： X_n 的增长速率 (或量级) 不超过 a_n 。换句话说， $\frac{X_n}{a_n}$ 是概率有界的 (bounded in probability)，即不会以显著概率变得过大。例子：如果 $X_n \sim N(0, 1)$ (标准正态分布)，则 $X_n = O_p(1)$ ，因为 X_n 的绝对值以高概率被某个常数 M 限制。

o_p (小 o 概率符号) 一个随机变量序列 X_n 满足 $X_n = o_p(a_n)$ ，如果对于任意 $\epsilon > 0$ ，当 $n \rightarrow \infty$ 时：

$$P\left(\left|\frac{X_n}{a_n}\right| > \epsilon\right) \rightarrow 0.$$

直观理解： X_n 的增长速率比 a_n 小得多，以至于 $\frac{X_n}{a_n} \rightarrow 0$ 在概率上。换句话说， X_n 在概率上比 a_n 更小量级。例子：如果 $X_n \sim N(0, \frac{1}{n})$ ，则 $X_n = o_p(1)$ ，因为 $\frac{X_n}{1} \rightarrow 0$ 在概率上。

2. 关系与区别

包含关系：如果 $X_n = o_p(a_n)$ ，则 $X_n = O_p(a_n)$ ，但反之不一定成立。

例如： $X_n = o_p(1)$ 意味着 $X_n \rightarrow 0$ 在概率上 (概率收敛)，而 $X_n = O_p(1)$ 只要求 X_n 概率有界。

特殊情况：

$X_n = o_p(1) \implies X_n \xrightarrow{p} 0$ (概率收敛到 0)。 $X_n = O_p(1) \implies X_n$ 是概率有界的。

3. 基本性质

加法:

$X_n = O_p(a_n)$, $Y_n = O_p(b_n)$, 则 $X_n + Y_n = O_p(\max(a_n, b_n))$ 。 $X_n = o_p(a_n)$, $Y_n = o_p(b_n)$, 则 $X_n + Y_n = o_p(\max(a_n, b_n))$ 。

乘法:

$X_n = O_p(a_n)$, $Y_n = O_p(b_n)$, 则 $X_n Y_n = O_p(a_n b_n)$ 。 $X_n = o_p(a_n)$, $Y_n = O_p(b_n)$, 则 $X_n Y_n = o_p(a_n b_n)$ 。

常数:

如果 c 是常数, 则 $cX_n = O_p(a_n)$ 如果 $X_n = O_p(a_n)$ 。 如果 $X_n = o_p(a_n)$, 则 $cX_n = o_p(a_n)$ 。

3.4 Donsker 定理 (Donsker's Theorem)

Donsker 定理是概率论和随机过程中的一个核心结果, 也称为函数空间中心极限定理 (Functional Central Limit Theorem, FCLT), 它描述了随机游走 (random walk) 在适当缩放下的极限行为趋向于布朗运动 (Brownian motion)。这是中心极限定理 (CLT) 在函数空间上的推广。1 定义与背景

随机游走: 给定独立同分布 (i.i.d.) 随机变量序列 $\{X_i\}_{i=1}^{\infty}$, 满足 $E(X_i) = 0$, $\text{Var}(X_i) = \sigma^2 < \infty$ 。定义部分和 (partial sum):

$$S_n = \sum_{i=1}^n X_i.$$

随机游走可以看作 S_n 随 n 的变化。目标: 研究 S_n 在时间轴上的行为, 特别是在适当的标准化后, 其路径是否收敛到一个随机过程。

2 Donsker 定理陈述设 $\{X_i\}$ 是 i.i.d. 随机变量序列, 满足 $E(X_i) = 0$, $\text{Var}(X_i) = \sigma^2$ 。定义标准化随机游走:

$$W_n(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}}, \quad t \in [0, 1],$$

其中 $[nt]$ 是小于等于 nt 的最大整数, $S_{[nt]} = \sum_{i=1}^{[nt]} X_i$ 。Donsker 定理表明:

$$W_n(t) \xrightarrow{d} W(t), \quad \text{as } n \rightarrow \infty,$$

其中:

\xrightarrow{d} 表示在函数空间 $D[0, 1]$ (由右连续且左极限存在的函数构成, 配备 Skorohod 拓扑) 上的分布收敛。 $W(t)$ 是标准布朗运动 (Wiener process), 满足 $W(0) = 0$, $E(W(t)) = 0$, $\text{Var}(W(t)) = t$ 。

3 直观解释

布朗运动: 布朗运动 $W(t)$ 是一个连续时间随机过程, 具有独立增量、正态分布 ($W(t) - W(s) \sim N(0, t - s)$), 路径几乎处处连续。Donsker 定理的意义: 离散的随机游走

路径 $W_n(t)$, 在适当缩放后 (时间轴缩放到 $[0, 1]$, 幅度缩放到 $\frac{1}{\sqrt{n}}$), 其分布会趋向于布朗运动。这将离散随机过程的极限行为与连续随机过程连接起来。

4 关键性质

函数空间: 收敛发生在 $D[0, 1]$ 空间, 而不是有限维空间 (如中心极限定理中的 \mathbb{R})。
 缩放因子: 时间缩放到 $\frac{[nt]}{n} \approx t$, 幅度缩放到 $\frac{1}{\sqrt{n}}$, 这是中心极限定理的直接结果。应用: Donsker 定理广泛用于统计学 (如时间序列分析、假设检验) 和金融数学 (如期权定价的极限行为)。

5 简单示例

假设 $X_i \sim \text{Bernoulli}(\pm 1)$, 即 $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$, 则 $E(X_i) = 0, \text{Var}(X_i) = 1$ 。
 定义:

$$W_n(t) = \frac{S_{[nt]}}{\sqrt{n}}.$$

当 $n \rightarrow \infty$, 路径 $W_n(t)$ 在 $D[0, 1]$ 中收敛到标准布朗运动 $W(t)$, 其方差为 $\text{Var}(W(t)) = t$ 。

3.5 伊藤积分 (Itô Integral)

伊藤积分是随机分析中的核心概念, 用于定义布朗运动驱动的随机过程的积分。它扩展了经典的黎曼积分, 适用于随机过程 (如布朗运动) 的非平滑路径。1 定义与背景

布朗运动: 标准布朗运动 $W(t)$ 是一个连续时间随机过程, 满足:

$W(0) = 0$, 增量 $W(t) - W(s) \sim N(0, t - s)$ (对于 $t > s$), 增量独立: 对于不重叠的时间区间, 增量相互独立。路径连续但几乎处处不可导。

目标: 定义 $\int_0^t f(s)dW(s)$, 其中 $f(s)$ 是一个随机过程 (例如依赖于 $W(s)$), $dW(s)$ 是布朗运动的增量。

2 伊藤积分的定义

对于一个适应随机过程 $f(t)$ (满足某些可积条件, 例如 $f(t) \in L^2$, 即 $E \left[\int_0^t f(s)^2 ds \right] < \infty$), 伊藤积分定义为:

$$\int_0^t f(s)dW(s) = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(t_{i-1})(W(t_i) - W(t_{i-1})),$$

其中:

$0 = t_0 < t_1 < \dots < t_n = t$ 是时间区间 $[0, t]$ 的一个划分, 划分的网格宽度 $\max_i(t_i - t_{i-1}) \rightarrow 0$ 。关键点: 积分中 f 的取值点选择为左端点 $f(t_{i-1})$, 这与经典黎曼积分不同, 体现了随机积分的“非预见性” (non-anticipating) 性质。极限: 上述极限是均方意义 (mean-square sense) 上的极限, 即:

$$E \left[\left(\int_0^t f(s)dW(s) - \sum_{i=1}^n f(t_{i-1})(W(t_i) - W(t_{i-1})) \right)^2 \right] \rightarrow 0.$$

3 伊藤积分的性质

线性：对于常数 a, b ，随机过程 f, g ：

$$\int_0^t (af(s) + bg(s))dW(s) = a \int_0^t f(s)dW(s) + b \int_0^t g(s)dW(s).$$

期望：如果 $f(s)$ 是适定的（例如 $f(s)$ 是可预测的且 $E \left[\int_0^t f(s)^2 ds \right] < \infty$ ），则：

$$E \left[\int_0^t f(s)dW(s) \right] = 0.$$

伊藤等距（Itô Isometry）：一个关键性质，用于计算方差：

$$E \left[\left(\int_0^t f(s)dW(s) \right)^2 \right] = E \left[\int_0^t f(s)^2 ds \right].$$

鞅性质：如果 $f(s)$ 满足适定条件，则 $\int_0^t f(s)dW(s)$ 是一个鞅（martingale），即：

$$E \left[\int_0^t f(s)dW(s) \middle| \mathcal{F}_s \right] = \int_0^s f(u)dW(u),$$

其中 \mathcal{F}_s 是布朗运动到时间 s 的信息集（滤波）。

4 简单推导：伊藤等距性质以 $f(s) = 1$ 为例，计算：

$$\int_0^t 1 dW(s) = W(t) - W(0) = W(t).$$

期望： $E[W(t)] = 0$ ，符合性质。方差：

$$E \left[\left(\int_0^t 1 dW(s) \right)^2 \right] = E[W(t)^2] = t,$$

$$E \left[\int_0^t 1^2 ds \right] = \int_0^t 1 ds = t.$$

两者相等，验证了伊藤等距性质。

5 应用：随机微分方程（SDE）伊藤积分是随机微分方程的基础。例如，考虑：

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t,$$

积分形式为：

$$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s.$$

其中第二项是确定性积分，第一项是伊藤积分，用于描述布朗运动驱动的随机波动。

6 伊藤引理（Itô's Lemma）

伊藤积分的一个重要应用是伊藤引理。设 X_t 满足 SDE:

$$dX_t = \mu_t dt + \sigma_t dW_t,$$

对于函数 $f(t, X_t)$, 伊藤引理给出:

$$df(t, X_t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dX_t)^2,$$

其中 $(dX_t)^2 = \sigma_t^2 dt$ (因为 $(dW_t)^2 \approx dt$)。展开后:

$$df(t, X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2} \sigma_t^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma_t \frac{\partial f}{\partial x} dW_t.$$