

基于 矩方法移动估计算法 的非平稳时间序列自适应 Student-t 分布

Jarek Duda

Jagiellonian University, Institute of Computer science and Computational Mathematics,
Łojasiewicza 6, Krakow, Poland, Email: dudajar@gmail.com

Abstract—现实生活中的时间序列通常是非平稳的，带来了模型适应性的难题。经典的方法如 ARMA-ARCH 假设任意类型的依赖关系。为了避免它们的偏差，我们将专注于最近提出的 agnostic 移动估计器哲学：在时间 t 找到优化例如 $F_t = \sum_{\tau < t} (1 - \eta)^{t-\tau} \ln(\rho_\theta(x_\tau))$ 移动对数似然参数，该参数随时间演变。这允许使用廉价的指数移动平均 (EMA) 来估计参数，如绝对中心矩 $m_p = E[|x - \mu|^p]$ 对于一个或多个幂 $p \in \mathbb{R}^+$ 的演变使用 $m_{p,t+1} = m_{p,t} + \eta(|x_t - \mu_t|^p - m_{p,t})$ 。将展示这种通用自适应矩方法在 Student's t 分布中的应用，该分布特别流行于经济应用中，此处应用于道琼斯工业平均指数公司的对数收益率。虽然标准的 ARMA-ARCH 方法提供了 μ 和 σ 的演化，但在这里我们还得到了描述 $\rho(x) \sim |x|^{-\nu-1}$ 尾部形状、极端事件概率的 ν 的演化——这些极端事件可能会导致灾难，破坏市场稳定。

关键词：非平稳时间序列，学生 t 分布，自适应模型，矩方法，重尾

I. 介绍

选择一个概率分布的参数族，例如这里的 Student's t-分布，通常关注的是直观上**静态估计**：整个数据集的一组参数 θ 的优化，通常是通过最大化某个评估指标如 $F = \frac{1}{T} \sum_{t=1}^T f(\theta, x_t)$ 。例如，在流行的 MLE (最大似然估计) 中使用 $f(\theta, x) = \ln(\rho_\theta(x))$ ，其中 $\rho_\theta(x)$ 是假设的参数族的概率分布函数 (PDF)。这样所有数据点都有相等的 $1/T$ 贡献，这似乎是对平稳时间序列的一个完美选择。

相比之下，现实生活中的时间序列往往是非平稳的，建议使用**自适应估计** [1]——具有演变参数，如图 1 中 Student's t-分布的 $\theta_t = (\mu_t, \sigma_t, \nu_t)$ 我们将重点关注。每个时间点的移动估计器 t 将基于前值 $\{x_\tau\}_{\tau < t}$ 分别优化 θ_t 参数，权重逐渐减弱，最终优化：

$$F = \frac{1}{T} \sum_{t=1}^T f(\theta_t, x_t) \quad \text{e.g. log-likelihood: } \frac{1}{T} \sum_{t=1}^T \ln(\rho_{\theta_t}(x_t)) \quad (1)$$

一个自然的方法来估计 θ_t 是通过优化类似的函数 F_t ：仅使

用过去值 $\{x_\tau\}_{\tau < t}$ ，并以指数递减的权重获得局部行为：

$$\theta_t = \arg \max_{\theta} F_t \quad \text{for} \quad F_t = \sum_{\tau < t} \bar{\eta}^{t-\tau} \ln(\rho_\theta(x_\tau)) \quad (2)$$

对于 $\bar{\eta} \in (0, 1)$ 学习率通常大于 0.9，也定义了 $\eta = 1 - \bar{\eta}$ 以便于计算。

上述 (2) 移动 MLE 可以很容易地直接优化 EPD (指数幂分布) 的 σ 尺度参数 $\rho(x) \sim \exp(-|x|^\kappa)$ [1] 瘦尾族，例如高斯和拉普拉斯分布，来自**绝对中心矩**: $m_p = E[|x - \mu|^p]$ ，用于适应性进化与指数移动平均 (EMA)：

$$m_{p,t+1} = m_{p,t} + \eta(|x_t - \mu_t|^p - m_{p,t}) \quad (3)$$

使用 $p = \kappa$ 进行 EPD，并将 μ_t 作为常量或也通过 EMA 进行调整。在这里我们将它应用到学生 t 分布上，这一次不是直接通过最大似然估计 (MLE)，因为没有显式公式，而是通过矩法来估计——从单次幂的绝对中心矩中估算出 σ 尺度参数，或者从两次幂的这样的矩中估算出 ν 自由度。

在一个关于道琼斯工业平均指数 (DJIA) 107 年的每日对数收益以及其最近 29 家公司的 10 年数据的例子中，测试了这种自适应估计方法，特别是对于 σ 的估计，这导致了实质上更好的对数似然评估，在这里，相对于 EPD [1]，学生 t 分布略好一些。这也比标准的 σ 预测方法如 GARCH(1,1) [2] 更优——从一方面来说，GARCH(1,1) 集中在高斯分布上，并且还任意假设了依赖关系——在这里被替换为一种不带偏见的移动估计器优化局部参数的方法。

这种自适应估计可以与其他方法结合，这些方法可能会在本文后续版本中添加。例如，在线 PCA [3] 或自适应线性回归 [4] 来结合来自多个来源的信息，比如这里的公司数据或宏观经济数据——以改进此处使用的参数估计的时刻预测。最后，如在 [1] 中讨论的那样，我们可以使用这种参数分布进行标准化 $y_t = \text{CDF}_t(x_t)$ ，然后用 HCR (分层相关重构) [5] 建模密度作为线性组合的方式，以静态或自适应 (随时间演变) 的方式来轻微扭曲均匀分布 $\{y_t\}$ 。

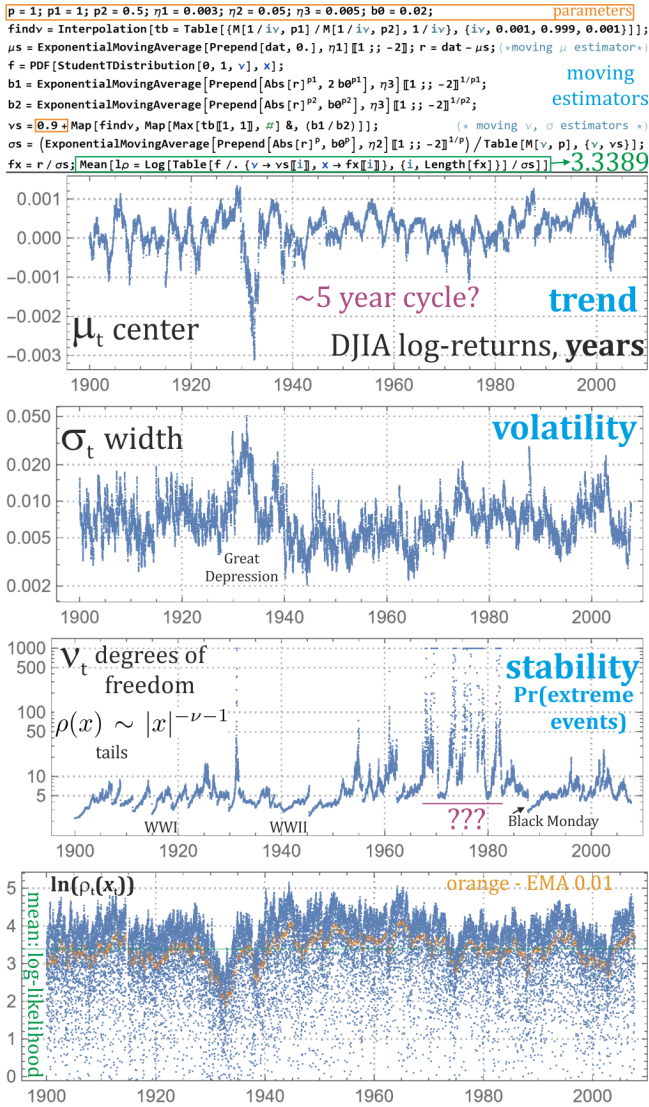


Figure 1. 用于移动估计所有 $\theta = (\mu, \sigma, \nu)$ Student's t 分布参数的 Mathematica 代码 (使用 $M_{\nu p} = E[|(x - \mu)/\sigma|^p]$ 矩公式 (6)), 及其应用于道琼斯工业平均指数 (DJIA) 107 年每日对数收益率时间序列的结果。这些参数在这种情况下是手动调整以最大化对数似然性的: 均值 $\ln(\rho_t(x_t))$ 显示在底部。我们可以看到这个世纪中有趣的演变, 这些可能值得进一步调查, 比如中心 μ 的 ≈ 5 年周期行为, 宽度 σ 的巨大 $\approx 25\times$ 变化, 以及主要在 1967 年至 1983 年间出现的一些几乎高斯分布的 $\nu \rightarrow \infty$ 时期。虽然 μ 描述了总体的上升/下降趋势, σ 接近波动性, 额外的 ν 以一种稳定性——潜在灾难性极端事件的概率来补充它。

II. 用于评估的时间序列

使用了 1900-2007 年道琼斯指数的每日数据¹, 基于每日对数收益率序列 $x_t = \ln(v_{t+1}/v_t)$ 进行工作。

图 2 还包括了 2018 年 9 月用于该指数的 30 家公司中 29 家公司的对数收益率评估。除了陶氏杜邦 (DWD) 之外, 过去十年的日价格数据均从纳斯达克网站 (www.nasdaq.com) 下载得到。对于其余的 29 家公司: 3M (MMM)、美国运通 (AXP)、苹果公司 (AAPL)、波音公司 (BA)、卡特彼勒公

¹DJIA 时间序列的数据来源: <http://www.idvbook.com/teaching-aid/data-sets/the-dow-jones-industrial-average-data-set/>

司 (CAT)、雪佛龙公司 (CVX)、思科系统公司 (CSCO)、可口可乐公司 (KO)、埃克森美孚公司 (XOM)、高盛集团 (GS)、家得宝公司 (HD)、IBM 公司 (IBM)、英特尔公司 (INTC)、强生公司 (JNJ)、摩根大通公司 (JPM)、麦当劳公司 (MCD)、默克公司 (MRK)、微软公司 (MSFT)、耐克公司 (NKE)、辉瑞公司 (PFE)、宝洁公司 (PG)、旅行者集团 (TRV)、联合健康集团 (UNH)、联合技术公司 (UTX)、威瑞森电信公司 (VZ)、维萨公司 (V)、沃尔玛公司 (WMT)、沃博联联盟公司 (WBA) 和华特迪士尼公司 (DIS), 均使用了从 2008 年 8 月 14 日到 2018 年 8 月 14 日期间的每日收盘价 (2518 个值)。

III. 学生 t 分布和适应性

学生 t 分布最早由弗里德里希·赫尔姆特在 1875 年引入 [6], 后来在 1908 年由威廉·西利·戈塞特以“学生”为笔名发表 [7], 从而得名。

其基本应用在于对 $\nu + 1$ 独立同分布高斯随机变量之和的分布: 对于样本均值与真实均值之间的差异。对于 $\nu = 1$ 来说是柯西分布, 对于大的 $\nu \rightarrow \infty$ 它逼近高斯分布。

其 PDF (概率密度函数), 如图 3 所示, 为:

$$\rho_{\mu\sigma\nu}(x) \equiv \rho(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{(x-\mu)^2}{\sigma^2\nu}\right)^{-\frac{1+\nu}{2}} \quad (4)$$

对于 $\mu \in \mathbb{R}$ 和 $\sigma, \nu \in \mathbb{R}^+$, $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ 伽玛函数。关键的是, 它具有多项式重尾 $\rho(x) \sim |x|^{-\nu-1}$ 对于 $|x| \rightarrow \infty$, 因此只有对于 $p < \nu$ 的有限矩 $E[x^p]$ 。

其累积分布函数 $\mu = 0, \sigma = 1$ 如下所示, 对于一般情况, 代入 $x \rightarrow (x - \mu)/\sigma$:

$$\int_{-\infty}^x \rho_{01\nu}(y)dy = \frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \frac{F_{1,2}\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}, -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \quad (5)$$

用于 $F_{1,2}$ 超几何函数。

A. 绝对中心矩方法

对于矩方法我们将使用绝对中心矩: $E[|x - \mu|^p]$ 用于不一定为整数的幂 $p \in \mathbb{R}^+$ 。使用 Mathematica 计算了如下积分的矩公式, 对于 $p < \nu$ 是有限的:

$$M_{\nu p} = \int_{-\infty}^{\infty} |x|^p \rho_{01\nu}(x)dx = \int_{-\infty}^{\infty} \frac{\nu^{p/2}\Gamma(\frac{p+1}{2})\Gamma(\frac{\nu-p}{2})}{\sqrt{\pi}\Gamma(\nu/2)} \quad (6)$$

具有 $\{x_t\}_{t=1..T}$ 数据样本, 固定 ν 并使用某些 μ 估计量, 例如将 $\hat{\mu} = T^{-1} \sum_t x_t$ 近似为均值, 上述公式给出了尺度参数 σ 的简单估计量:

$$\hat{\sigma} = \frac{\sqrt{T^{-1} \sum_t |x_t - \hat{\mu}|^p}}{M_{\nu p}} \quad (7)$$

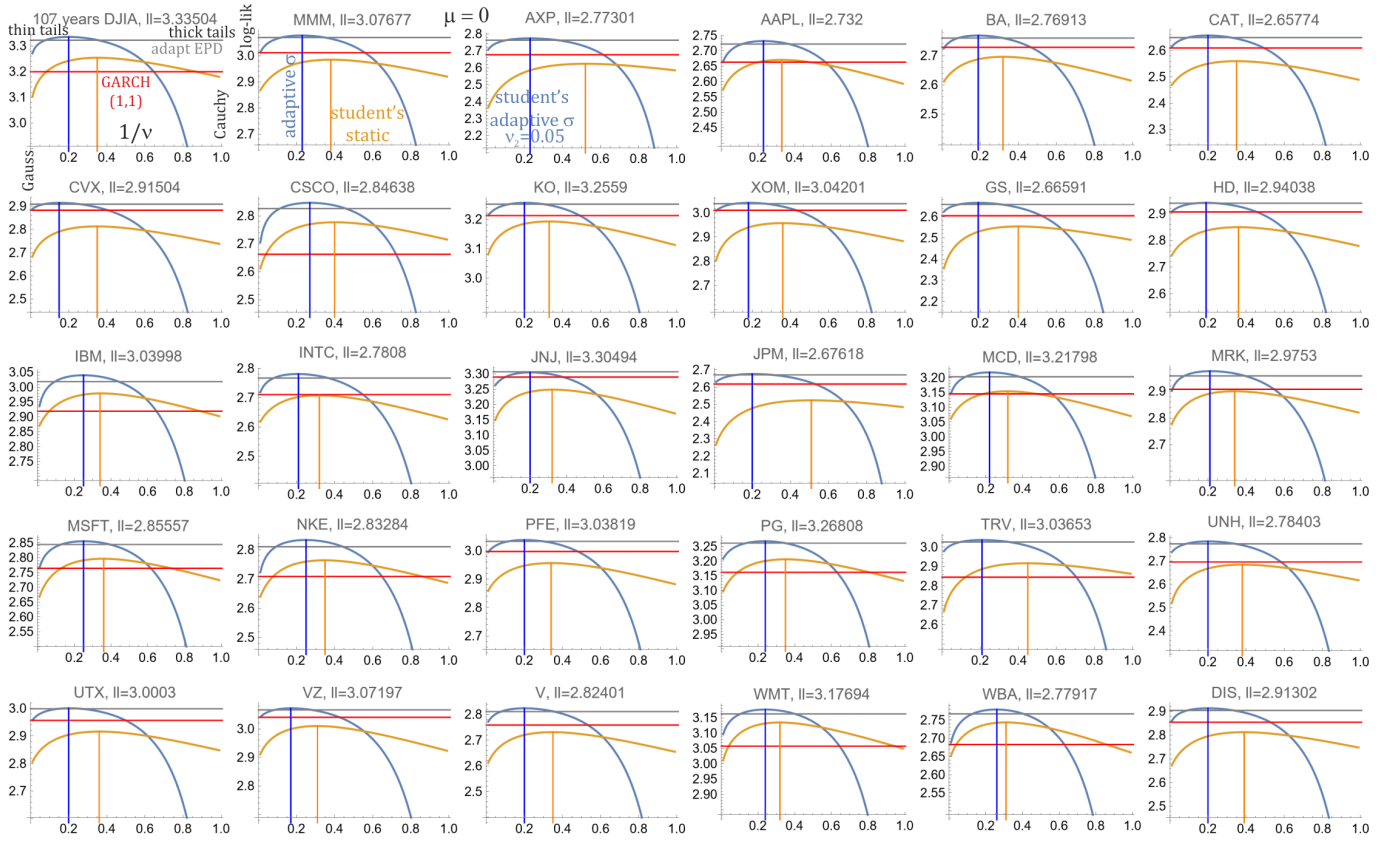


Figure 2. 对数似然值（均值 $\ln(\rho_t(x_t))$ ）评估了 107 年道琼斯工业平均指数时间序列的对数收益率，以及 29 家单独公司的 10 年的数据。在横轴上，有 $1/\nu$ 学生 t 分布的自由度（从高斯分布到柯西分布），对于静态参数（橙色）和自适应 σ 尺度参数（蓝色，使用 $p = 1$ 次幂和 $\eta_2 = 0.05$ 学习率），所有这些都是在 $\mu = 0$ 中心。我们可以看到，适应性使得尾部更轻（最大值时 ν 更大）。还类似地展示了从 σ 适应的最优结果用于 $\rho(x) \sim \exp(-|x|^\kappa)$ 指数幂分布，如前文 [1]（灰色）所示。红线显示了通过标准 GARCH(1,1) 模型评估的 σ 适应情况——这与 $\nu = \infty$ 高斯情形可比，但通常略差一些。

所使用的 p 必须在 $(0, \nu)$ 范围内，其中使用非整数 p 可能对 $p < \nu$ 要求至关重要。

此外，使用各种 p 进行此类 σ 估计存在不同的不确定性，这取决于 ν ，如图 4 所示——建议优化 p ，例如基于所使用的 ν 范围，甚至动态修改 p 。对于较大的 ν ，最优的 p 接近于 $p = 2$ 方差估计，标准为 $\nu \rightarrow \infty$ 高斯分布极限。对于较小的 ν ，最优的 p 是 $\approx \nu/6$ 。

为了估计 ν ，一个自然直接的方法是将两个不同幂次的平均值 p_1, p_2 进行划分，去除 σ 依赖性：

$$\frac{M_{\nu p_1}}{M_{\nu p_2}} \approx \frac{p_1 \sqrt{N^{-1} \sum_i |x_i - \hat{\mu}|^{p_1}}}{p_2 \sqrt{N^{-1} \sum_i |x_i - \hat{\mu}|^{p_2}}} \quad (8)$$

选择一些 $p_1 \neq p_2$ ，则 $M_{\nu p_1}/M_{\nu p_2}$ 随 ν 单调变化（示例见图 4），我们可以例如将其行为放入表格中，并基于平均值进行插值得到 ν 的估计，例如在图 1 的代码中以 `find $\nu$` 方式实现。

然而，类似于 $1/(n-1)$ 的方差估计量的标准调整，(8) 的估计似乎有偏差——需要通过计算其期望值来进行调整，最好是用一个明确的公式（尚待找到）。在图 1 中，这样的轻微调整是通过对找到的 ν 加上（调节过的）0.9 来实现的。

B. 移动中心矩估计量

上述矩法可以通过简单地将平均值替换为指数移动平均值——均匀权重替换为指数减弱的权重来轻松适应移动估计器。

对于中心 μ ，我们可以使用下面的基本适应性调整——它仅在高斯情形（ $\nu \rightarrow \infty$ ）下是最佳的，因此通常情况下可以稍作改进。然而，对于所讨论的数据来说，收益已经几乎可以忽略不计了。

$$\mu_{t+1} = \mu_t + \eta_1(x_t - \mu_t) \quad (9)$$

最为关键的是 σ 尺度参数自适应估计，例如在 ARCH 家族中，但以更不可知的方式进行，在这里使用 (7) 公式为选定的 $p \in (0, 2)$ 次幂（ $p < \min_t(\nu_t)$ ），这次使用随时间演变的（中心绝对）矩：

$$m_{p,t+1} = m_{p,t} + \eta_2(|x_t - \mu_t|^p - m_{p,t}) \quad (10)$$

最后，对于 ν 自由度的估计，我们可以使用 (8) 公式来更新某些不同幂次 p_1, p_2 的矩，并且采用某种 η_3 学习率。

图 1 包含了所有 3 个参数适应所使用的 Mathematica 代码，以及它们在 DJIA 时间序列中的演化。手动调整导致了 3 种不同的学习率： $\eta_1 = 0.003, \eta_2 = 0.05, \eta_3 = 0.005$ 分

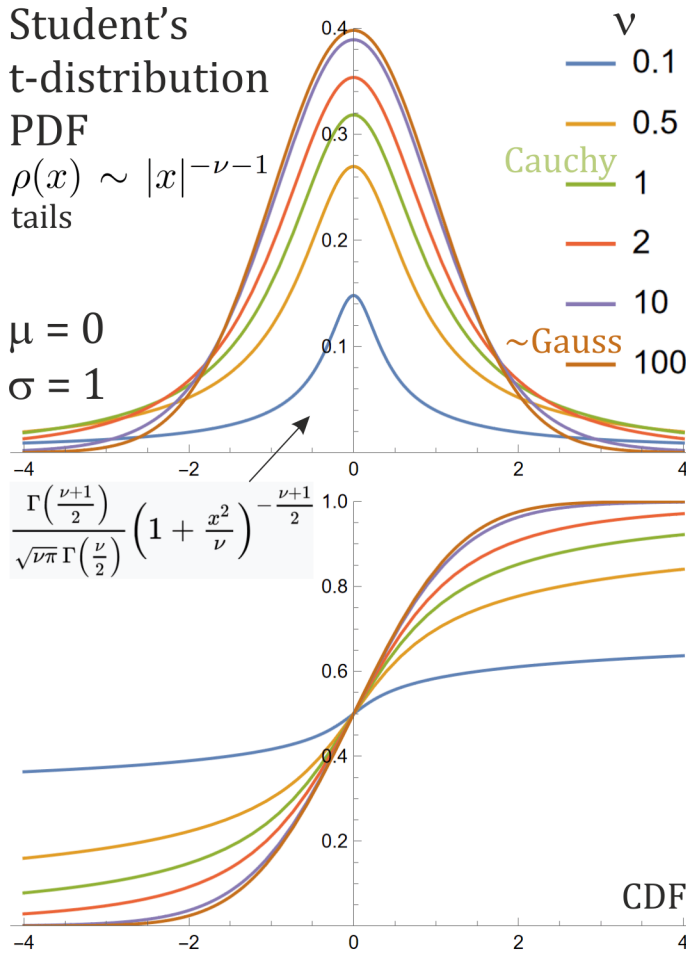


Figure 3. 概率分布函数 (PDF, 渐近于 $\sim |x|^{-1-\nu}$) 和累积分布函数 (CDF) 对于中心固定为 $\mu = 0$ 和尺度参数为 $\sigma = 1$ 的学生 t 分布, 但形状参数不同为 ν 。当 $\nu \rightarrow \infty$ 时我们得到高斯分布, $\nu = 1$ 时得到柯西分布, 并且还可以涵盖不同类型重尾和分布体。

别对应于 μ, σ, ν (对于尺度参数 σ 快得多)。

图 2 显示了使用固定 $\mu = 0$ 中心和各种固定 ν 对单个 MLE 参数 σ 进行评估, 或者使用 (7) 估计的 σ 调整, 具有 $p = 1$ 次幂和 $\eta_2 = 0.05$ 学习率。例如: 导致对数似然仅比图 1 中优化演化所有 3 个参数时差 ≈ 0.004 。 ν 估计器需要调整——此处通过简单添加调优参数来完成, 希望未来能改进并自动化。

ν 进化, 在标准的 ARMA-ARCH 方法中不可用, 评估局部尾形, 潜在破坏性极端事件的概率——建议将其称为**稳定性**, 补充流行的**波动率**评估, 类似于 σ 。图 5 和 6 检查了确实如图 1 所建议的 1967-1983 范围具有更薄的尾部。图 7 显示了所有公司 ($\eta_3 = 0.005$) 的 ν 演变情况——这种分析可能有助于定位和理解影响稳定性的因素/机制, 这有望在未来应用于降低潜在灾难性极端事件发生的概率。

C. 包括不对称性

尤其是尾部, 例如图 6 所示, 明显是非对称的, 因此将其纳入可能也是有价值的。有两种经典的方法可以不对称

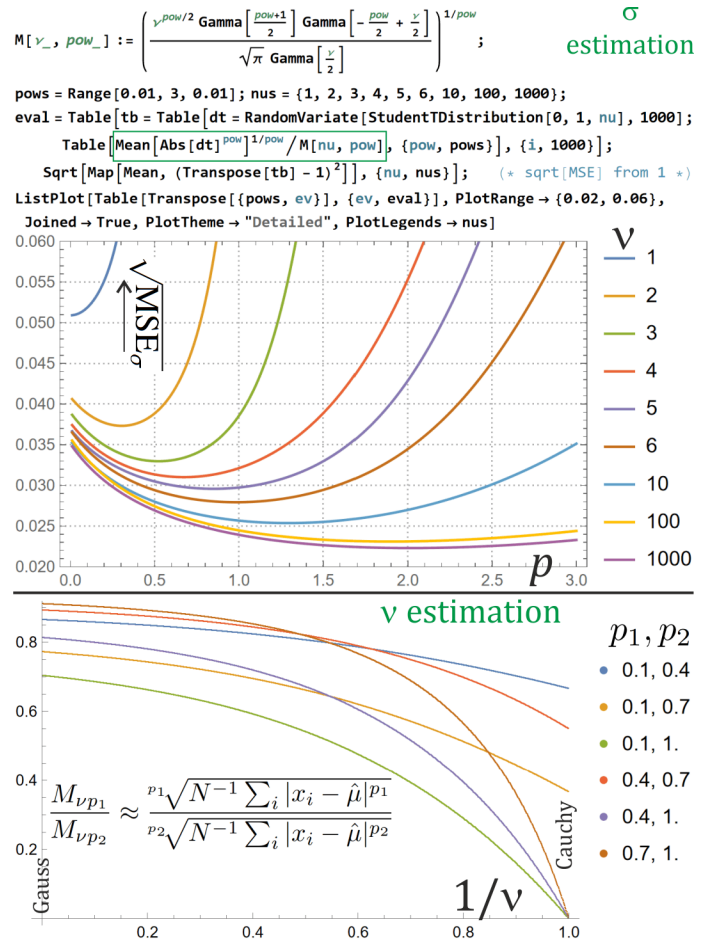


Figure 4. 顶部: 错误依赖于在 σ 估计中选择的幂 p 作为 $\hat{\sigma} = \sqrt{N^{-1} \sum_i |x_i - \hat{\mu}|^p / M_{\nu p}}$ 。我们可以看到, 对于高斯分布 $\nu \rightarrow \infty$, 我们应该选择 $p = 2$, 就像在标准方差估计中一样, 但为了提高预测应该减少这个 p 对于较低的 ν 到 $p \approx \nu/6$ 。底部: 单调函数用于估计 ν 的各种 2 次幂选择 p_1, p_2 。

Numbers of extreme events for 107 years DJIA daily log-returns: data vs expected Student's t-distribution

event	$\sigma = \sqrt{\text{var}}$	adapt	$\nu=1$ Cauchy	$\nu=2$	$\nu=3$	$\nu=4$	$\nu=5$	$\nu=6$	$\nu=10$	$\nu=\infty$ Gauss
1σ	5204	11191	14674.5	12404.3	11475.5	10973.6	10660.1	10445.8	10004.9	9312.75
2σ	1170	3281	8662.86	5385.64	4089.08	3407.9	2991.82	2712.62	2153.87	1335.39
3σ	422	916	6011.64	2801.83	1692.52	1172.26	883.383	704.617	391.623	79.2363
4σ	194	316	4577.22	1678.5	822.02	473.402	302.982	208.935	73.9105	1.85904
5σ	96	133	3688.17	1107.91	451.753	219.837	120.469	71.9738	15.7702	0.0168259
6σ	58	74	3085.66	782.781	272.145	113.949	54.1823	28.3081	3.87726	0.0008579107
7σ	43	48	2651.23	581.226	175.691	64.3368	26.9056	12.4288	1.09049	7.51224 × 10 ⁻⁶
8σ	29	37	2323.47	448.103	119.643	38.8551	14.4663	5.97148	0.345583	3.65158 × 10 ⁻¹¹
9σ	28	29	2067.54	355.759	84.9892	24.7656	8.29637	3.08961	0.121448	6.62459 × 10 ⁻¹⁵
10σ	23	20	1862.22	289.16	62.4664	16.4942	5.01714	1.69989	0.0466518	4.4727 × 10 ⁻¹⁹

Above all 29349 days, below 4012 days 1967-1983 - no 6σ (extreme) events, much closer to Gaussian

event	1967-1983	$\nu=1$ Cauchy	$\nu=2$	$\nu=3$	$\nu=4$	$\nu=5$	$\nu=6$	$\nu=10$	$\nu=\infty$ Gauss
1σ	1078	2006.	1695.67	1568.7	1500.09	1457.23	1427.94	1367.66	1273.05
2σ	204	1184.21	736.216	558.976	465.859	408.981	370.814	294.433	182.547
3σ	38	821.789	383.009	231.368	160.247	120.758	96.3209	53.5347	10.8316
4σ	13	625.705	229.45	112.37	64.7139	41.4175	28.5613	10.1036	0.25413
5σ	3	504.172	151.451	61.7545	30.0516	16.4681	9.8388	2.15578	0.00230009

Figure 5. 实际和预期的事件数量 $|X - \mu| > k\sigma$: 对于 $k = 1, \dots, 10$, 完整的 29349 个值的时间序列从 1900-2007 (顶部) 以及限制在 4012 个值上从 1967-1983 (底部)。显著的数据中的数值数量, 在右侧有各种 ν 的 Student t 分布的预期事件数量 (长度 × 概率)。在顶部表格中, 我们看到大量的极端事件, 在使用接近 $\nu \in (3, 5)$ 的自适应 σ 后。相比之下, 由图 1 中 ν 的演化所建议的 1967-1983 年区间具有极低的 $\nu \sim 10$ 极端事件概率——这表明市场更加稳定。图 7 显示了更详细的 ν 演化, 可能有助于定位、理解关键机制, 并且也许可以利用这些机制来使市场更加稳定。

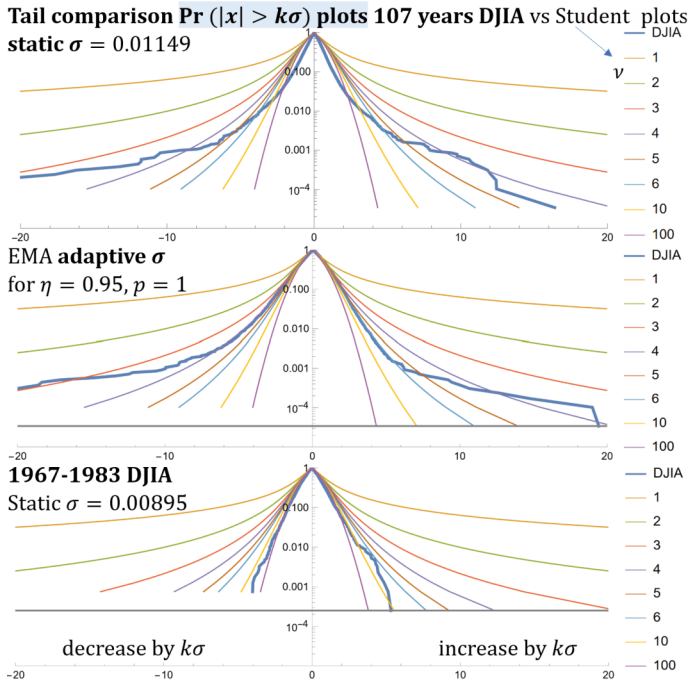


Figure 6. 图5的可视化表格：超出 $k\sigma$ 的概率向左（负）和右（正）：基于道琼斯工业平均指数数据（粗蓝色线），并与各种自由度 ν 的学生 t 分布进行比较（细色线）。顶部：整个 107 年期间的静态 $\sigma = 0.001149$ ，我们可以观察到中心行为在对数尺度下几乎呈线性，如同拉普拉斯分布。中心自适应 σ 使中心行为更接近学生 t 分布，但尾部对应于 3 到 6 之间的各种 ν ，左右尾之间有明显的不对称。底部：如所见，1967-1983 年的道琼斯工业平均指数几乎呈高斯分布，这可以在限制在此期间的数据的底部图中看到，已经为静态 σ 提供了良好的一致性，并且尾部在 10 到 100 之间的 ν 。

化 Student-t 分布：非中心 t 分布 [8] 的概率密度函数为

$$\frac{e^{-\frac{\delta^2}{2}} 2^\nu \nu^{\frac{\nu}{2}+1} \Gamma\left(\frac{\nu+1}{2}\right)}{\pi} H_{-\nu-1}\left(\frac{-x\delta}{\sqrt{2}\sqrt{x^2+\nu}}\right) (\nu+x^2)^{-\frac{\nu+1}{2}}$$

和参数 ν, δ ：和偏斜广义 t 分布 [9] 的概率密度函数。

$$\frac{p}{2v\sigma q^{\frac{1}{p}} B\left(\frac{1}{p}, q\right) \left(1 + \frac{|x-\mu+m|^p}{q(v\sigma)^p(1+\lambda \operatorname{sgn}(x-\mu+m))^p}\right)^{\frac{1}{p}+q}}$$

然而，它们复杂得多，尤其是如果想要寻找自适应估计，这可能是未来值得考虑的内容。

相反，为了简化起见，让我们从不对称性开始，只需将两个具有独立参数的标准 Student t 密度粘合到 μ 中，这些参数描述了 $\overleftarrow{\sigma}, \overleftarrow{\nu}$ 左尾，以及描述了 $x < \mu, \overrightarrow{\sigma}, \overrightarrow{\nu}$ 右尾。虽然这不是必要的，但为了使概率密度函数连续，我们需要确保在 μ 处左部和右部具有相同的值，通过线性缩放这会导致概率密度函数：

$$\rho_{\mu \overleftarrow{\sigma} \overleftarrow{\nu} \overrightarrow{\sigma} \overrightarrow{\nu}}(x) = \frac{2 \begin{cases} \left(1 + \frac{(x-\mu)^2}{\overleftarrow{\sigma}^2 \overleftarrow{\nu}}\right)^{-\frac{\overleftarrow{\nu}+1}{2}} & \text{if } x \leq \mu \\ \left(1 + \frac{(x-\mu)^2}{\overrightarrow{\sigma}^2 \overrightarrow{\nu}}\right)^{-\frac{\overrightarrow{\nu}+1}{2}} & \text{if } x > \mu \end{cases}}{\sigma_l \sqrt{\overleftarrow{\nu}} B\left(\frac{\overleftarrow{\nu}}{2}, \frac{1}{2}\right) + \sigma_r \sqrt{\overrightarrow{\nu}} B\left(\frac{\overrightarrow{\nu}}{2}, \frac{1}{2}\right)} \quad (11)$$

这样做的优点是我们可以像之前一样精确地使用自适

应估计 σ, ν ，只是将当前 μ 以下/以上的值分开：我们保留两份 σ, ν 的副本，并根据 $\operatorname{sgn}(x - \mu)$ 更新其中一个。

然而，对 DJIA 的测试导致了分别适应左右尾部的 $\overleftarrow{\nu}, \overrightarrow{\nu}$ 和 $\overleftarrow{\sigma}, \overrightarrow{\sigma}$ 的 ≈ 0.01 更差的近似。它回到了分别适应 $\overleftarrow{\nu}, \overrightarrow{\nu}$ 和共同 $\sigma = \overleftarrow{\sigma} = \overrightarrow{\sigma}$ 的原始近似。因此，虽然 $\overleftarrow{\nu}, \overrightarrow{\nu}$ 可以用来描述单独左右尾部的形状，但对于 σ 尺度参数来说，使用一个共同的是更好的选择。我们还可以寻找更复杂的自适应估计方法，或者可能使用非中心或偏斜 t 分布。

我们可以使用这种演变特别是 ν 来评估市场：例如，分别估计左尾和右尾极端事件的概率，也将它们用作各种模型的局部参数。图 8 显示了它在 107 年道琼斯工业平均指数中的演化过程，有趣的是，显示出某个时期正好有一个尾部较重的情况：无论是减少还是增加都能更容易地获得极值。

IV. 结论与进一步工作

本文介绍了矩方法的一些新颖扩展——不仅适用于幂次不一定是自然数的绝对中心矩（在处理低 ν 时至关重要），更重要的是作为 EMA 移动估计器——用于随时间变化的参数，也可以独立地应用于左右尾。除了能够更好地评估对数似然之外，它还能提供这些关键参数的变化情况，如图 1、7、8 所示——包括 ν 自由度来评估极端事件的概率，理解这种依赖关系可能会允许引入一些市场稳定机制。例如，它建议寻找在 1967 年至 1983 年间道琼斯工业平均指数中 ν 急剧增加的机制，在图 5、6 中得到确认。

这是一个通用方法，可能也值得应用于其他分布，如 alpha-稳定分布和更大的模型。同时，将其与其它特别是自适应模型结合也是有意义的，比如在线性回归和 HCR（层次相关重构）——这将在本文后续版本中进行规划。

进一步工作的计划示例：

- 改进来自矩的估计量——特别是 ν 的。
- 改进不对称情况的演化，例如更好地适应性估计所提出的粘合的两个学生 t 分布，或非中心、偏斜变体。
- 增加进一步的建模，例如依赖其他股票、宏观经济数据等，比如使用自适应线性回归 [4] 和 HCR [5] 来包括细微的相关性。
- 寻找各种分布的移动估计器的不同方法，例如使用梯度上升方法，可能还包括如 [10] 中的二阶信息。
- 讨论的方法有许多超参数，比如学习率——通常对于相似的数据类型是通用的。自动优化它们并通过进化进行适应可能是有价值的。
- 理解影响 ν 演化的机制/依赖关系，也分别针对左尾和右尾，并希望利用它们，例如提高标记的稳定性。
- 测试了适用于不同应用的方法，如数据压缩，其中对数似然改进转化为每符号节省的比特。

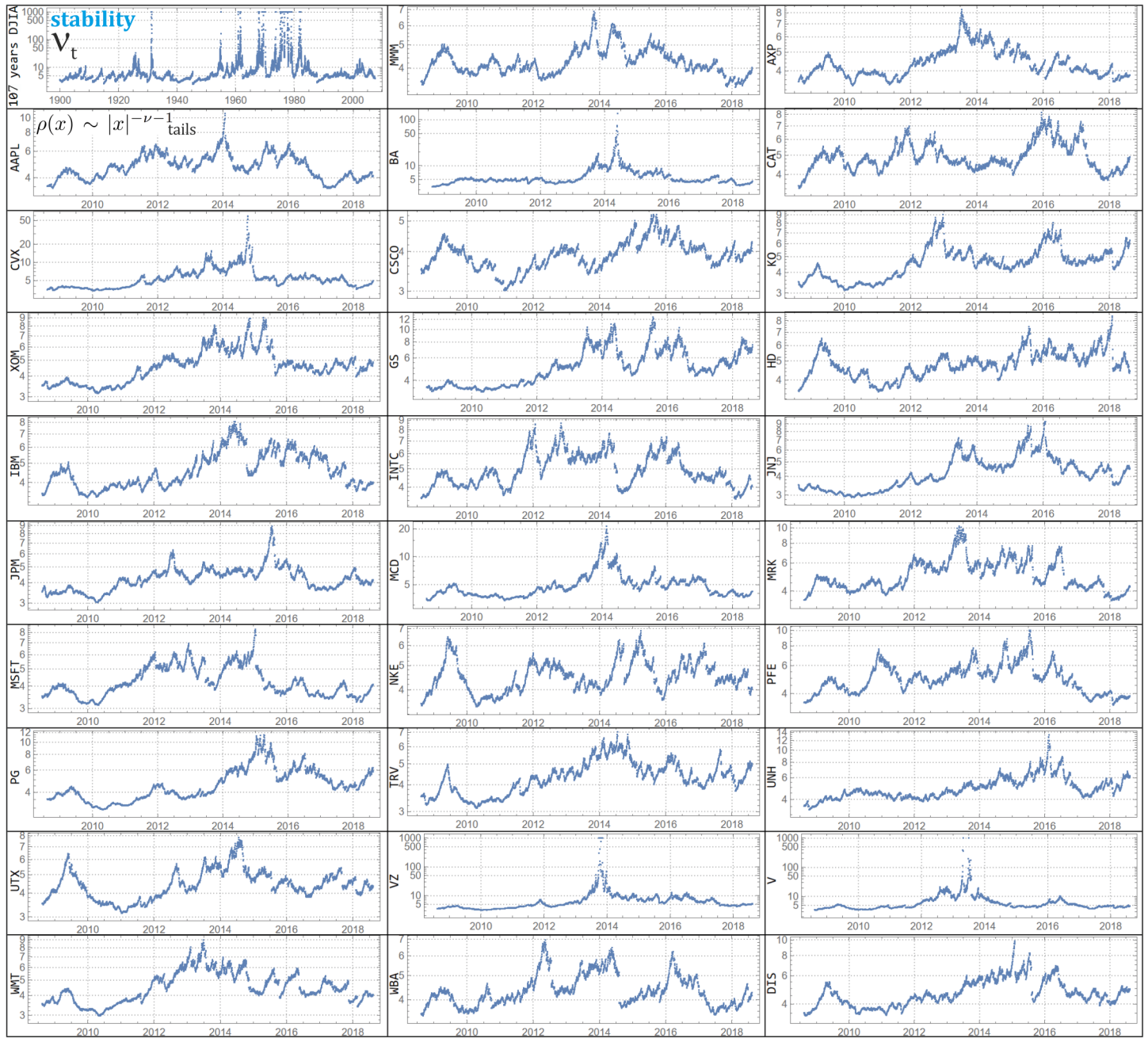


Figure 7. ν 参数在所有 1+29 个案例中的演变，使用了 $p_1 = 1$ 和 $p_2 = 1/2$ 次方以及 $\eta_3 = 0.005$ 学习率。它描述了尾部形状 $\rho(x) \sim |x|^{-\nu-1}$ ，极端事件的概率——可能是灾难性的，可能会导致市场不稳定，这表明了“稳定性”解释补充了标准的“波动性”评估。将上述演变与各种历史事件/因素进行比较可能有助于理解和利用它们来改善市场稳定。

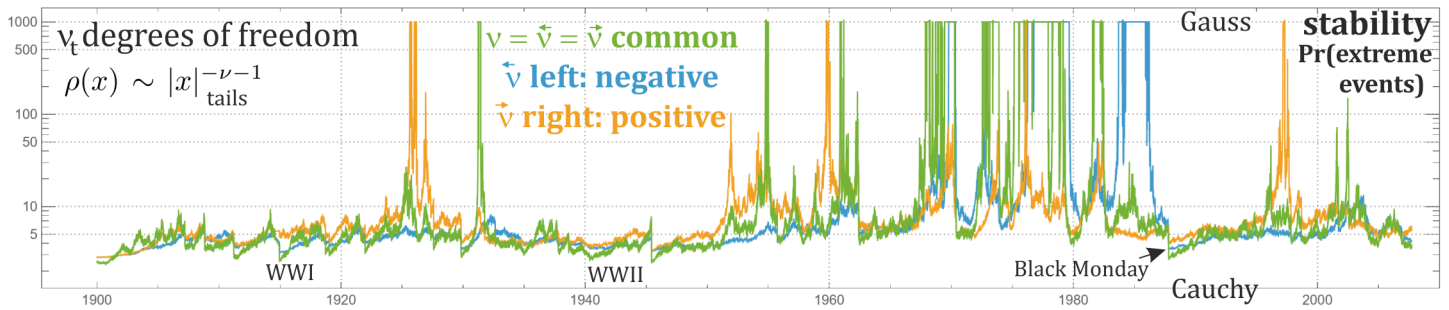


Figure 8. 如图 6 所示存在尾部不对称，因此进行了如下分析：对道琼斯工业平均指数 (DJIA) 的自由度 ν 进行自适应估计，如同之前在分布 1 中所做的那样（绿色），并且单独分析了负值左侧尾部 $x_t < \mu_t$ 的 $\hat{\nu}$ （蓝色）和正值右侧尾部 $x_t > \mu_t$ 的 $\hat{\nu}$ （橙色）。有趣的是，我们可以观察到如 1983-87 年间仅有单侧重尾的现象。类似的左右单独估计 σ 更加嘈杂，并且仅恶化了对数似然，因此没有展示。

REFERENCES

- [1] J. Duda, “Adaptive exponential power distribution with moving estimator for nonstationary time series,” *arXiv preprint arXiv:2003.02149*, 2020.
- [2] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [3] H. Cardot and D. Degras, “Online principal component analysis in high dimension: Which algorithm to choose?” *International Statistical Review*, vol. 86, no. 1, pp. 29–50, 2018.
- [4] J. Duda, “Parametric context adaptive laplace distribution for multimedia compression,” *arXiv preprint arXiv:1906.03238*, 2019.
- [5] —, “Exploiting statistical dependencies of time series with hierarchical correlation reconstruction,” *arXiv preprint arXiv:1807.04119*, 2018.
- [6] F. R. Helmert, “Über die berechnung des wahrscheinlichen fehlers aus einer endlichen anzahl wahrer beobachtungsfehler,” *Z. Math. U. Physik*, vol. 20, no. 1875, pp. 300–303, 1875.
- [7] Student, “The probable error of a mean,” *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.
- [8] N. Johnson and B. Welch, “Applications of the non-central t-distribution,” *Biometrika*, vol. 31, no. 3/4, pp. 362–389, 1940.
- [9] P. Theodossiou, “Financial data and the skewed generalized t distribution,” *Management science*, vol. 44, no. 12-part-1, pp. 1650–1661, 1998.
- [10] J. Duda, “Improving SGD convergence by tracing multiple promising directions and estimating distance to minimum,” *arXiv preprint arXiv:1901.11457*, 2019.