# Learning Market Regimes

Marcos Costa Santos Carreira

XP Inc

QuantMinds - London, 19-Nov-2024

# Contents

# Standard causality applies

Acknowledgements

- XP Inc
- Emanuel Derman

If this work is good, it's because of them; if it's not good, it's not their fault

# What

- One needs to know what a regime is before defining how it could change
- So it must have some regularity
- And some weak form of the anthropic principle (the change cannot be destructive enough to wipe out everything, there should be some kind of recovery)
- So it should have some kind of cycle

# When

- Earliest description of regime changes
- Although one could say the Flood, we'll focus on Joseph's interpretation of the Pharaoh's dream (Genesis 41)
- "Behold, there come seven years of great plenty throughout all the land of Egypt. And there shall arise after them seven years of famine; and all the plenty shall be forgotten in the land of Egypt; and the famine shall consume the land; ... and take up the fifth part of the land of Egypt in the seven years of plenty. And let them gather all the food of these good years that come, and lay up corn under the hand of Pharaoh for food in the cities, and let them keep it. And the food shall be for a store to the land against the seven years of famine, which shall be in the land of Egypt; that the land perish not through the famine."

# Why

- We can distinguish here all of our goals:
  - Marked changes from the previous situations
  - An eventual return to "normality"
- We also see that prediction is not enough
  - "Do not spend the wealth from the years of plenty" means:
  - Your actions are a consequence of the past and a function of the possibilities of the future (RL)
- This work is inspired by Derman's classic "Regimes of Volatility"

# How

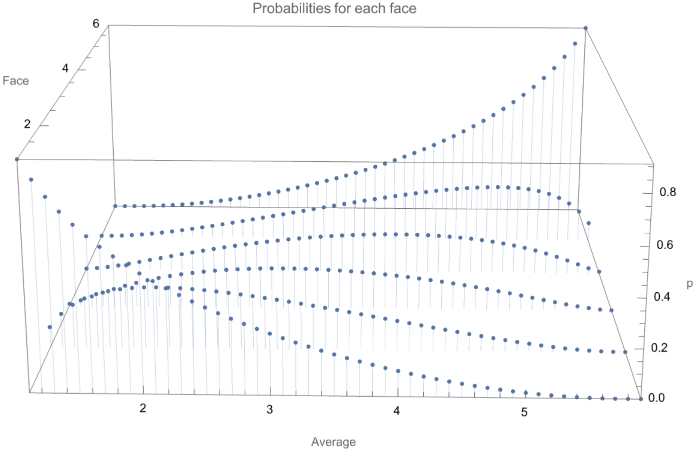
- We start from an insurance-like perspective
  - "I just know that something ~~good~~ bad is gonna happen / I don't know when"
  - Maximum Entropy
- Then move to the idea of regime switching
  - "Leonid Will Tear Us Apart"
  - But what are we losing?
  - "I Can See Clearly Now"
- Are there other ways to mix the history with the recent past?
  - "Memory, Uncaring Friend"
  - Rough Volatility Regressions
  - Interest Rates and Central Bank Decisions
- Can we model paths?

# Roll The Bones

- Maximum Entropy Distributions are those that will keep the maximum uncertainty about a random variable given what we know with certainty (typically a moment of some function of the random variable)
- An example is Jaynes' analysis of the Brandeis Dice problem
- We know nothing particular about a dice => we should expect an uniform distribution as the probability of each face
- But if we say that after a really large number of tosses the average number of spots was 4.5 (instead of 3.5), what can we conclude?

# Horror vacui

- The answer is not the the probabilities of faces 1 and 2 are equal to 0; instead, the probabilities should be increasing with the number of spots but never zero for any face:



# How Can I Be Sure?

- It takes a lot to say something is impossible
- Although a Gaussian is the Maximum Entropy Distribution given the mean and the variance of a sample, knowing the variance is quite a strong claim
- (This is why we're here after all)
- Can we use the Student's T Distribution to understand what do we need to go from the nihilism of the Cauchy to the certainty of the Gaussian?

# Finally some equations

What do we (think we) know?

- The expected value of $\ln\left(1+\frac{x^2}{\nu}\right)$:

$$\int_{-\infty}^{+\infty}\left[p\left(x\right)\cdot\ln\left(1+\frac{x^2}{\nu}\right)\right]\mathrm{d}x=\omega \tag{1}$$

What do we get?

- Something similar to the normalized Student's T PDF:

$$p\left(x\right)=\frac{\left(1+\frac{x^2}{\nu}\right)^{\theta(\omega)}}{\sqrt{\nu}\cdot B\left(-\theta\left(\omega\right)-\frac{1}{2},\frac{1}{2}\right)}=\frac{\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}}{\sqrt{\nu}\cdot B\left(\frac{\nu}{2},\frac{1}{2}\right)} \tag{2}$$

So we want to match:

$$\nu=-2\cdot\theta\left(\omega\right)-1 \tag{3}$$

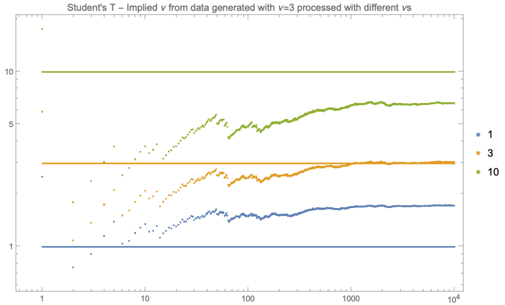
With the inverse function $\theta$ defined as ($\psi$ is the Digamma function):

$$\theta\left(\omega\right)=x\ ,\ \left[\psi\left(-x\right)-\psi\left(-x-\frac{1}{2}\right)\right]=\omega$$

# If it doesn't fit you must ~~acquit~~ try another n

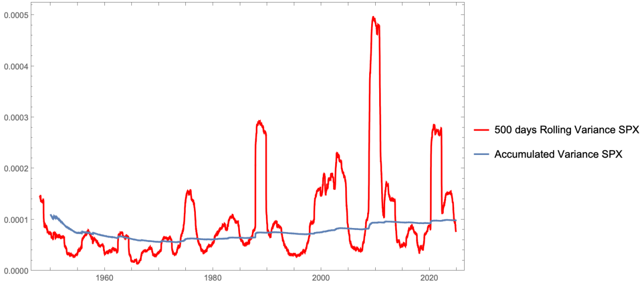
So if the equation $\nu = -2 \cdot \theta(\omega) - 1$ doesn't fit, ie the result from $\nu = -2 \cdot \theta(\omega) - 1$ derived from transforming the data with $\ln\left(1 + \frac{x^2}{\nu}\right)$ is not coherent with the $\nu$ we used, we must try another value of $\nu$ to transform the data.

An example with data generated from a Student's T with $\nu = 3$ (calculating the accumulated mean of the data with 3 different $\ln\left(1 + \frac{x^2}{\nu}\right)$ transformations and calculating $\nu = -2 \cdot \theta(\omega) - 1$):



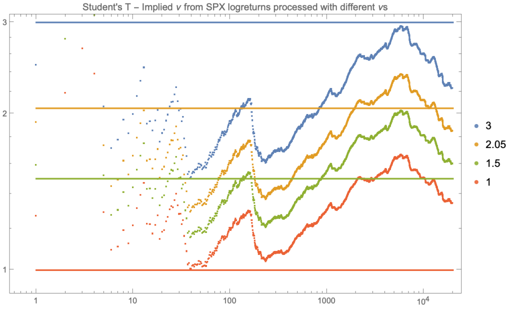
# Voluble volatilities

We think the Gaussian would not be good (SPX example):



# Say goodbye to the variance

And even the Student's T is barely viable; if you want a fit that allows for a variance we need $\nu = 2.05$ (or something similar just above 2), and even then there's no sign of a convergence after decades of data ($\nu = 1.5$ looks better):



# Regime Switching

- A common solution for voluble volatilities is regime switching, where 2 (or more) regimes characterized by different volatilities (or different distributions) switch according to transition matrices.
- Example: An index is modeled as a low vol (12%) regime and a high vol (30%) regime, with transition probabilities 10% for low to high and 25% high to low.
- It might be interesting to have a tool to detect a regime change as early as possible
    - A classic paper is Adams and MacKay ("Bayesian Online Changepoint Detection")

# Leonid Will Tear Us Apart

- One of the common metrics is the p-Wasserstein Distance (recent papers by Blanka Horvath and others):
  - For $p = 1$ we have:

$$W_1 = \int_{-\infty}^{+\infty} \left| F_1(x) - F_2(x) \right| \mathrm{d}x \tag{4}$$

- This is a distance between distributions
  - Let's see some examples

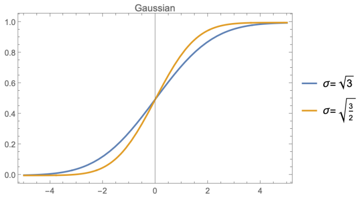
# Just sigma

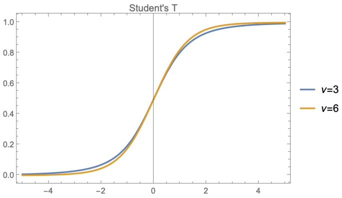
One characteristic of the Gaussian is that it can be summarized through its sufficient statistics (mean and variance)

- The Wasserstein distance between two centered Gaussian (mean=0) is:

$$W_1(\sigma_1, \sigma_2) = |\sigma_1 - \sigma_2| \cdot \sqrt{\frac{2}{\pi}} \tag{5}$$

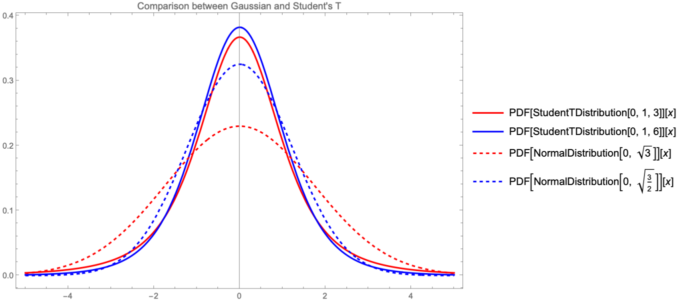
But the distance between the two centered standardized Student's T distributions with the same variances as the Gaussians is smaller



# How different you are

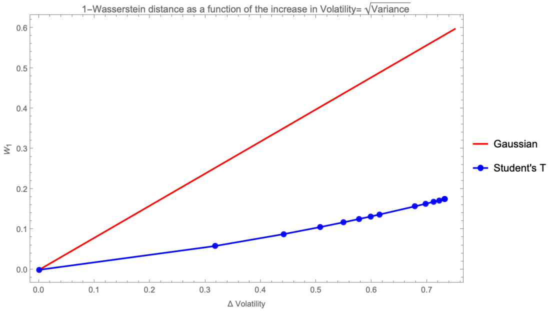
- We can see it clearly on the PDFs




- The fact that the Gaussians are distinguishable means they are not MaxEnt distributions - they need to be clearly separable and this is much easier without fat tails.

# A small step

If we plot the change in the square root of the Variance for the two distributions we can see how the Gaussians are more distant for the same change in variance:



# To infinity and beyond!

- But what's the problem with using Gaussians?
- The Gaussian distribution is really lousy for Expected Shortfall, as there is an illusion that any new record will not exceed the previous record by much
- So the main danger is not necessarily using a mixture of Gaussians to identify regime changes, but to believe that the future realizations can be modeled through a mixture of Gaussians
- The general lesson is that if you want a clear distinction between distributions it'll be hard to make them fat-tailed.

# Durations

- In previous presentations on microstructure we discussed how volatility could be seen less as a frequentist statistic about the number of price changes within a period and more of a GBM parameter with direct influence on the expected time between price changes
- This allows us to use a probability distribution to model and update volatility (with all the caveats of how hard it is to distinguish between $\mu$ and $\sigma$
- And these times between price changes follow a lognormal distribution as expected from their relationship with the volatilities
- This is important - a market with a predictable time window for a price change seems too predictable

# I Can See Clearly Now

- But this is exactly what happens in crashes and regime changes
- The predictive power of Order Book Imbalance increases, and the predictive power of the past few moves increases as well
- It's this reduction of uncertainty, when both informed and noise traders have an alignment of information and reach the same conclusions that is the hallmark of regime changes
- Remember how the Gaussian was characterized by the sufficient statistics - after a few realizations the information brought by a new realization goes to zero
- It's when each new realization is very informative (ie it has predictive power over the next realizations) that we have a regime change

## Rough Vol Regressions

- There's a great paper by Mikko Pakkanen and others ("Decoupling the short and long-term behavior of stochastic volatility") which we used to investigate the coefficients of the log-variance prediction
- From:

$$\mathbb{E}\left[\log\left(\sigma^2_{t+\Delta}\right)|\mathcal{F}_t\right] \approx \frac{\cos\left(H\cdot\pi\right)}{\pi}\Delta^{H+\frac{1}{2}}\int_{-\infty}^{t}\frac{\log\left(\sigma^2_s\right)}{\left(t-s+\Delta\right)\left(t-s\right)^{H+\frac{1}{2}}}\mathrm{d}s \tag{6}$$

- We reach:

$$\begin{aligned}\mathbb{E}\left[\log\left(\sigma^2_{\Delta}\right)|\mathcal{F}_0\right] \approx & \frac{\cos\left(H\cdot\pi\right)}{2\pi}\left(\psi\left(\frac{3}{4}-\frac{H}{2}\right)-\psi\left(\frac{1}{4}-\frac{H}{2}\right)\right)\log\left(\sigma^2_0\right)\\ &+\sum_{j=2}^{\infty}\left\{\left(-1\right)^{-H-\frac{1}{2}}B_{(-j),(1-j)}\left(\frac{1}{2}-H,0\right)\log\left(\sigma^2_{(1-j)\cdot\Delta}\right)\right\}\end{aligned} \tag{7}$$

# Efficient Coefficients

- Where $\psi$ is the Digamma function and $B_{x_1,x_2}(a,b)$ is the incomplete beta function, defined as:

$$B_{x_1,x_2}(a,b) = \int_{x_1}^{x_2} x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

- Rewriting:

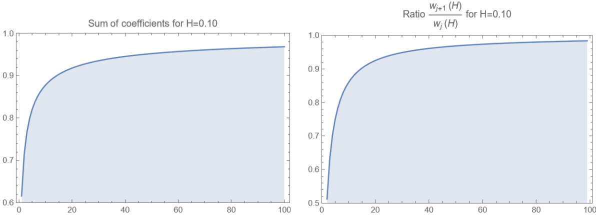
$$\mathbb{E}\left[\log\left(\sigma_\Delta^2\right)|\mathcal{F}_0\right] \approx w_1(H)\log\left(\sigma_0^2\right) + \sum_{j=2}^{\infty}\left\{w_j(H)\log\left(\sigma_{(1-j)\cdot\Delta}^2\right)\right\} \tag{8}$$

# H is enough

This formula has 4 noteworthy characteristics:
(i) The weights of the log-variances depend only on H
(ii) The 1st weight $w_1(H)$ is such that $\frac{1}{2} \leq w_1(H) \leq 1$ for $0 \leq H \leq \frac{1}{2}$; in fact, it can be approximated by $\frac{1}{2} + H$
(iii) The other weights $w_j(H)$ are products of two complex numbers, but the imaginary part of the product gets chopped (literally: when using Mathematica or Python's mpmath to deal with complex numbers the function to get rid of the small terms is Chop)
(iv) The sum of the weights converges very slowly to 1, and they decay really slowly (after 20 points one could almost treat them as the same number)

# Memory, Uncaring Friend

- If weight #499 is basically the same as weight #500 (and they are still needed):




- Then the order is not important, and our memory only cares about the frequency of these past realizations
- So we could replace the sum after 5 or 10 terms with the stationary distribution of log-variance

# Keeping the tails

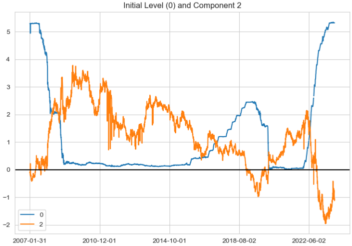
- So by being able to sample from the stationary distribution without bias (but with a low weight) and using the recent past in a regression we end up with something similar to Guyon's PDV model (without the trend component): a fast decay kernel and a long memory kernel, but without losing the ability to sample from the tails
- This blurs the boundaries among regimes, since it might be hard to have a clear demarcation between periods of low and high volatility

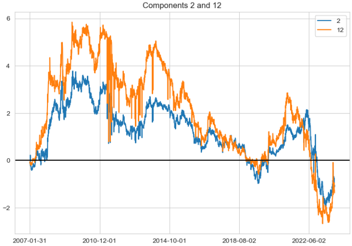
# Anchors

- When looking at the dynamics of Interest Rate Term Structures, it is important to remember that, in most countries guided by an inflation targeting regime, these variables are important:
  - The past few decisions of the Central Bank (ie, which kind of cycle we're in, if any)
  - And the level of the rate (there is some mean reversion, 10 year rates have had a narrower range than short rates)
- The episodes where the short rate was close to (or below) zero show how this affected the dynamics
- And Curvature is related to the Slope (papers and talks by Andreasen, Sokol, etc.)

# US Rates

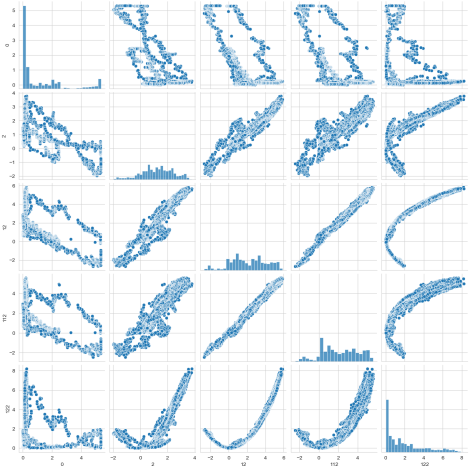
- Signature components of USD Rates:



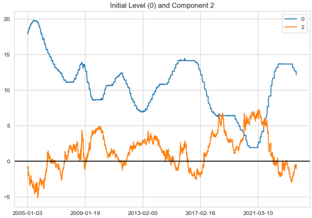
# US Rates

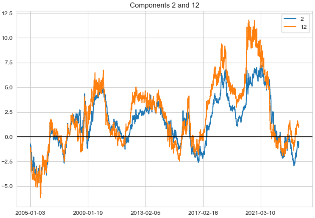
- Signature components of USD Rates:



# Brazilian Rates

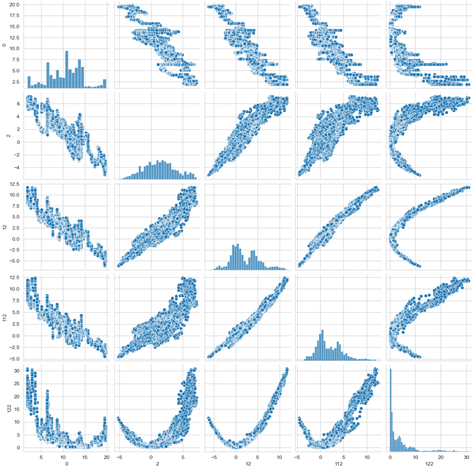
- Signature components of Brazilian Rates:



# Brazilian Rates

- Signature components of Brazilian Rates:



# Methods and code

- Port Mathematica code to Python
- Comparison with BOCD and Blanka's results
- Publish results

# Data and models

- Investigate where paths (as typical sequences, including order when useful, or as their signature representations) can work as components of regimes
    - Bouchaud
    - Blanka
- Multivariate
    - Maximum entropy copulas
    - Distances are harder to calculate

# What this talk was about anyway?

Regime changes

- We start without clustering or sequence, using maximum entropy
- "Sticky" distributions need to be distinguishable and therefore "narrow"
- Regime changes are characterized by an increase in predictive power
- IR curves have relatively few degrees of freedom, anchored by Central Bank cycles and some degree of mean reversion

Prediction

- Rough volatilities can be useful as a "fast" kernel and a complete history with fat tails
- Guyon's PDV is interesting but roughness is a feature not a bug
- Paths are promising

# Papers


- Carreira, M. C. S.. (2023), “Core Signatures and Inversions” Core Signatures and Inversions (other papers and notes are available at ResearchGate or my GitHub)
- Derman, E. (1999), “Regimes of Volatility - Some Observations on the Variation of S&P 500 Implied Volatilities” Regimes of Volatility
- Bennedsen, M., Lunde, A. and Pakkanen, M. (2016) “Decoupling the short- and long-term behavior of stochastic volatility” arXiv:1610.00332
- Horvath, B and Issa, Z. (2023) “Non-parametric online market regime detection and regime clustering for multidimensional and path-dependent data structures” SSRN:4493344
- Morel, R., Mallat, S. and Bouchaud, J-P. (2023). “Path Shadowing Monte-Carlo” arXiv:2308.01486
- Adams, R. P. and MacKay, D. J. C. (2007). “Bayesian Online Changepoint Detection” arXiv:0710.3742

# Books


- Kapur, J.N., "Maximum Entropy Models in Science and Engineering" New Age International Publishers (1989)
- Cover, T. M. and Thomas, J. A. "Elements of Information Theory" Wiley, 2nd Edition (2006)