Diego Alvarez Selected Works

Diego.alvarez@colorado.edu | GitHub | LinkedIn | Medium

Simplified approach for $dW_t dt = 0$, $dW_t^2 = dt$, and $dt^2 = 0$

Diego Alvarez
diego.alvarez@colorado.edu

July 2021

Introduction

When going over Ito's lemma it may be common to see the proof for the differentials, and not really fully understand where they come from. When I first learned Ito's lemma I remember glossing over the proof and really just remembering the common differentials.

This paper is more of how these computations are done, therefore this document is similar to the math in an appendix of a paper.

Overview of Steps

Something to consider is that each of these statements are really integrals and not differentials

$$dW_t dt = \int_0^t dW_s ds, \quad dW_t^2 = \int_0^t dW_s^2, \quad dt^2 = \int_0^t ds^2$$

The steps for proving these statements

1. Write out definition of the integral.
2. *pack* variables.
3. Then set up the integral to be proved by mean squared convergence
4. Work out mean squared convergence and then *unpack* variables
5. Use the fact that expectation is a linear operator and work out each expectation

Things we'll need later on

Because this paper consists of math that would normally be in the appendix, we will cover some of the proofs and tools that we will need in the future rather than referencing them during the proof causing the reader to go to the appendix.

These are some simple algebra proofs that we will use later on

$$(a+b)^2 = a^2 + b^2 + 2ab \tag{1}$$

$$(a-b)^2 = a^2 + b^2 - 2ab \tag{2}$$

For two independent Brownians

$$\begin{aligned} \mathbb{E}[\Delta W_{t_k} \Delta W_{t_j}] &= \mathbb{E}[\Delta W_{t_k}]\mathbb{E}[\Delta W_{t_j}] \\ &= \mathbb{E}[\Delta W_{t_j}]\mathbb{E}[\Delta W_{t_k}] \\ &= 0 \end{aligned} \tag{3}$$

If this statement true then the random variable X_n converges in the mean squared sense. This is mean-squared convergence

$$\mathbb{E}[|X_n - X|^2] = 0 \tag{4}$$

Let's write out the first moments of a Brownian motion B_t

$$E[B_t] = 0$$

$$E[B_t^2] = t$$

Working out $dW_t dt = 0$

Let's start by representing this as an integral

$$dW_t dt = \int_0^t dW_s dt$$

Then by definition of integral

$$\begin{aligned} \int_0^t dW_s dt &= \lim_{n\to\infty} \sum_{k=1}^{n} \left(W_{t_k} - W_{t_{k-1}}\right)(t_k - t_{k-1}) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \Delta W_{t_k} \Delta t_k \end{aligned} \tag{5}$$

Now we will start *packing* our variables by using these substitutions

$$\Delta W_{t_k} \Delta t_k = X_k \tag{6}$$

$$X_n = \sum_{i=1}^{n} X_k \tag{7}$$

Applying the substitutions in eq.6 and eq.7 into eq.5

$$\lim_{n\to\infty} X_n = X \tag{8}$$

Now we can plug in X_n into the mean-squared convergence from eq.4

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^2] = 0$$

We are checking that it goes to zero ($dW_t dt = 0$) so we set $X = 0$ giving us

$$\lim_{n\to\infty} \mathbb{E}[|X_n - 0|^2] = 0$$

Now let's start *unpack* our variables

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_k\right)^2\right] \tag{9}$$

Then by the proof in eq.1 we can rewrite the summation. The thing to consider is that a and b are values in an index. We have to take that into account when we work out the summation

$$\left(\sum_{k=1}^{n} X_k\right)^2 = \sum_{i=1}^{n} X_k^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} X_k X_j$$

Then plugging that into eq.9 we get

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_k\right)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} X_k^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} X_k X_j\right]$$

Then *unpack* using our substitutions in eq.6 and eq.7

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_k\right)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} (\Delta W_k \Delta t_k)^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} \Delta W_{t_k} \Delta t_k \Delta W_{t_j} \Delta t_j\right]$$

Now use the fact that expected value is a linear operator

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_k\right)^2\right] = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[(\Delta W_{t_k})^2\right] + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} \mathbb{E}\left[\Delta W_{t_k} \Delta W_{t_j}\right] \Delta t_k \Delta t_j$$

From eq. 3 we know that 2nd expected value $\mathbb{E}\left[\Delta W_{t_k} \Delta W_{t_j}\right]$ goes to 0. That leaves

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_k\right)^2\right] = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[(\Delta W_{t_k})^2\right]$$

Now we are left with $\mathbb{E}\left[(\Delta W_{t_k})^2\right]$ which is variance, and the second moment. The variance of a Brownian motion the length of the interval, but by the second moment from the moment generating function in eq. [add equation] is

$$\lim_{n\to\infty}\sum_{k=1}^{n}\Delta t_k \Delta t_k^2 = \lim_{n\to\infty}\sum_{k=1}^{n}\Delta t_k^3 = \lim_{n\to\infty}\sum_{k=1}^{n}(t_k - t_{k-1})^3$$

That all goes to zero, from a visual perspective we cutting an interval to an infinitesimal size.

$$\lim_{n\to\infty}\sum_{k=1}^{n}\Delta t_k^3 = 0$$

Now that the limit goes to zero we have proved that $dW_t dt = 0$ goes to by the mean squared convergence theorem

Working out $dW_t^2 = dt$

Let's start by representing this as an integral

$$dW_t^2 = \int_0^t dW_s^2$$

Then by definition of integral

$$\begin{aligned} \int_0^t dW_s^2 &= \lim_{n\to\infty}\sum_{k=1}^{n}\left(W_{t_k} - W_{t_{k-1}}\right)^2 \\ &= \lim_{n\to\infty}\sum_{k=1}^{n}\Delta W_{t_k}^2 \end{aligned} \tag{10}$$

In this case we are not going to *pack* and *unpack* variables, but we are going to summation into the mean squared convergence

$$\sum_{k=1}^{n}\Delta W_{t_k}^2 = \mathbb{E}\left[\left|\sum_{k=1}^{n}\Delta W_{t_k}^2 - X\right|^2\right] = 0$$

Because we are trying to prove that $dW_t^2 = dt$ we set $X = t$ and plugging that into the limit as well we get

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}\Delta W_{t_k}^2 - X\right|^2\right] = \lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{k=1}^{n}\Delta W_{t_k}^2 - t\right)^2\right] = 0 \tag{11}$$

Then by the proof in eq.2 we can rewrite the summation as

$$\left(\sum_{k=1}^{n}\Delta W_{t_k}^2 - t\right)^2 = \left(\sum_{k=1}^{n}\Delta W_{t_k}^2\right)^2 + t^2 - 2t\sum_{k=1}^{n}\Delta W_{t_k}^2$$

And plugging that into eq.11

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{k=1}^{n} \Delta W_{t_k}^2 - t\right)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{k=1}^{n} \Delta W_{t_k}^2\right)^2 + t^2 - 2t\sum_{k=1}^{n} \Delta W_{t_k}^2\right] \tag{12}$$

Then for the first summation on the right hand side we can use the proof from eq.12

$$\left(\sum_{k=1}^{n} \Delta W_{t_k}^2\right)^2 = \sum_{k=1}^{n} X_k^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{n-1} X_k X_j$$

Plugging all of that into the equation

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{k=1}^{n} \Delta W_{t_k}^2 - t\right)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\sum_{k=1}^{n} \Delta W_{t_k}^4 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} \Delta W_{t_k}^2 \Delta W_{t_j}^2 + t^2 - 2t\sum_{k=1}^{n}]\Delta W_{t_k}^2\right]$$

Then use the fact that expectation is a linear operator

$$\lim_{n\to\infty} \left(\sum_{k=1}^{n} \Delta\mathbb{E}\left[W_{t_k}^4\right] + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} \mathbb{E}\left[\Delta W_{t_k}^2 \Delta W_{t_j}^2\right] + t^2 - 2t\sum_{k=1}^{n} \mathbb{E}\left[\Delta W_{t_k}^2\right]\right)$$

Using the moment generating function that we found we can work out the moments

$$\mathbb{E}\left[\Delta W_t^4\right] = 3\Delta t^2$$

$$\mathbb{E}\left[\Delta W_t^2\right] = \Delta t$$

And for independent Brownians we can separate the expectations

$$\mathbb{E}\left[\Delta W_{t_k}^2 \Delta W_{t_j}^2\right] = \mathbb{E}[\Delta W_{t_k}^2]\,\mathbb{E}[\Delta W_{t_j}^2]$$

Plugging in expected values that we got

$$\lim_{n\to\infty} \left(3\sum_{k=1}^{n} \Delta t_k^2 + 2\sum_{k=1}^{n} \Delta t_k \Delta t_j + t^2 - 2t\sum_{k=1}^{n} \Delta t_k\right)$$

Then if we look at this summation

$$\sum_{k=1}^{n} \Delta t_k = t$$

Replacing that summation with t gives us

$$\lim_{n\to\infty} \left(3\sum_{k=1}^{n} \Delta t_k^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1} \Delta t_k \Delta t_j - t^2\right)$$

By definition we can rewrite

$$\Delta t = \frac{t}{n}$$

Plugging that in we get

$$\lim_{n\to\infty}\left(3\sum_{k=1}^{n}\left(\frac{t}{n}\right)^2 + 2\sum_{k=1}^{n}\sum_{j=1}^{k-1}\frac{t}{n}\cdot\frac{t}{n} - t^2\right)$$

Then we can rewrite the summations are

$$\sum_{k=1}^{n} = n, \quad \sum_{j=1}^{k-1} = \frac{n-1}{2}$$

So plugging that in gives us

$$\lim_{n\to\infty}\left(3n\left(\frac{t}{n}\right)^2 + 2\cdot\frac{n(n-1)}{2}\cdot\left(\frac{t}{n}\right)^2 - t^2\right)$$

Then simplify

$$\lim_{n\to\infty}\left(\frac{3t^2}{n} + \left(1-\frac{1}{n}\right)t^2 - t^2\right)$$

Now when we consider the limit the fractions go to 0 which leave

$$0 + (1+0)t^2 - t^2 = t^2 - t^2 = 0$$

We have now proved that $dW_t^2 = dt$

Working out $dt^2 = 0$

Let's represent the differential as an integral

$$dt^2 = \int_0^t ds^2$$

Then by definition of the integral

$$\int_0^t ds^2 = \lim_{n\to\infty}\sum_{k=1}^{n}\Delta t_k^2$$

Then we are trying to prove this statement by using mean squared convergence

$$\sum_{k=1}^{n}\Delta t_k^2 = \mathbb{E}\left[\left|X_n - X\right|^2\right] = 0$$

Then plug in our desired value $X = 0$

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\sum_{k=1}^{n} \Delta t_k^2 - 0\right|^2\right] = 0$$

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{k=1}^{n} \Delta t_k^2\right)^2\right] = 0$$

In this case everything is deterministic so we can drop the expected value

$$\lim_{n\to\infty} \left(\sum_{k=1}^{n} \Delta t_k^2\right)^2$$

Then use the same definition that we had used in the previous example

$$\Delta t = \frac{t}{n}$$

$$\sum_{k=1}^{n} = n$$

Then plugging all that in we get

$$\begin{aligned} \lim_{n\to\infty} \left(\sum_{k=1}^{n} \Delta t_k^2\right)^2 &= \lim_{n\to\infty} \left(n\left(\frac{t}{n}\right)^2\right)^2 \\ &= \lim_{n\to\infty} \left(\frac{t^2}{n}\right)^2 \end{aligned} \tag{13}$$

Then as we evaluate the limit the fraction goes to 0.

$$\lim_{n\to\infty} \left(\frac{t^2}{n}\right)^2 = 0$$

We have now shown that it goes to zero therefore the statement is true by mean squared convergence.

Diffusion Equation

Diego Alvarez

September 2021

Contents

1 Introduction

2 Initial Model

Let's first start with the function $f(x,t)$ and to model the changes 1-dimensional dynamics. If we increase time by a small τ we get

$$f(x, t+\tau) = \int_{-\infty}^{\infty} f(x+\Delta, t)\phi(\Delta)d\Delta \tag{1}$$

Where Δ is the change in displacement that the particle makes. The function $\phi(\Delta)$ is the probability distribution of that particle's path. Essentially we are translating the future position of a particle is really the distance it will travel in time multiplied by the probability of that path being taken.

The cornerstone to this theory was representing both functions using taylor series.

We want to find the the taylor series for $f(x+\Delta,t)$

$$f(x+\Delta,t) = f(x,t) + \Delta \cdot \frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!} \cdot \frac{\partial^2 f(x,t)}{\partial x^2} + ... + \frac{\Delta^n}{n!} \cdot \frac{\partial^n f(x,t)}{\partial x^n}$$

Then apply that to the definition of $f(x,t+\tau)$ that we had found in eq.1

$$f(x,t+\tau) = f(x,t)\int_{-\infty}^{\infty}\phi(\Delta)d\Delta + \frac{\partial f(x,t)}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta + \frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2!}\phi(\Delta)d\Delta + ... + \frac{\partial^n f(x,t)}{\partial x^n}\int_{-\infty}^{\infty}\frac{\Delta^n}{n!}\phi(\Delta)d\Delta$$

Let's look at the first term, we can see that the integral goes to 1.

$$f(x,t)\int_{-\infty}^{\infty}\phi(\Delta)d\Delta = f(x,t)\cdot 1$$

Then if we look at the the second term it goes to zero.

$$\frac{\partial f(x,t)}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta = 0$$

The best way to think of this is that when we integrate the Δ we will get an even exponent and when we plug in the bound we'll end up subtracting the same value. This is the case for all even terms in the our taylor series.

Now we want to find show that this all goes to the diffusion equation. We'll also need the taylor series of $f(x,t+\tau)$

$$f(x,t+\tau) = f(x,t) + \tau\cdot\frac{\partial f(x,t)}{\partial t} + \frac{\tau^2}{2!}\cdot\frac{\partial^2 f(x,t)}{\partial t^2} + ... + \frac{\tau^n}{n!}\cdot\frac{\partial^n f(x,t)}{\partial t^n}$$

Here is the slick trick with the taylor series that we found for $f(x,t+\tau)$ and the taylor series for $f(x+\Delta,t)$ that we applied to our definition. Really beyond the 2nd terms they partials don't yield much. The way I like to think of it is by thinking of what the actual derivatives mean. As we keep taking more partial derivatives with respect to time we get velocity → acceleration → jerk → snap → crackle → pop. The movements of particles don't really exhibit acceleration, and therefore don't exhibit the other derivatives of position vector.

Applying that gives us

$$f(x,t+\tau) = f(x,t) + \frac{\partial f(x,t)}{\partial t}$$

Although we'll keep the other higher moments for the partial derivatives with respect to x

$$f(x,t) + \frac{\partial f(x,t)}{\partial t} = f(x,t) + \frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\phi(\Delta)\frac{\Delta^2}{2!}d\Delta + \text{higher moments}$$

Now get rid of the $f(x,t)$ and divided by τ

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2\tau}\cdot\phi(\Delta)d\Delta$$

Then if we let D be the mass diffusivity we get

$$\frac{\partial f(x,t)}{\partial t} = D\cdot\frac{\partial^2 f(x,t)}{\partial x^2}$$

3 Solution to Diffusion Equation

3.1 Difference Method

Let's start with a general diffusion equation

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2} \tag{1}$$

The goal is to find an invariant transformation that reduces the order of the PDE

$$v = \lambda x$$

$$u = \lambda^2 t$$

We know that the transformation is invariant (see Appendix A and B). In this case we will add on a constant to our transformation making the function $af(\lambda x, \lambda^2 t)$. We know that when we integrate our function we will end up with the number of particles.

$$\int_{mathbbR} af(\lambda x, \lambda^2 t)dx = n$$

The trick here is to switch integration. Let's recall that $v = \lambda x$ to switch integration to dv we need to take the derivative. Another way to think of it is by doing u-substitution but in this case our dummy variable is v. (See appendix C.)

$$\frac{a}{\lambda}\int_{\mathbb{R}} f(v, \lambda^2 t)dv = n$$

We know that this statement is true (integrating over the function gives us the number of particles)

$$\int_{\mathbb{R}} f(v, \lambda^2 t)dv = n$$

Therefore to make that statement true with the $\frac{a}{\lambda}$ we have to set that fraction equal to 1. Then from there we can solve for λ

$$\frac{a}{\lambda} = 1$$

Now the question becomes what should we set λ to. Let's look again at the diffusion equation which is

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2}$$

The left hand side represents

$$\frac{\partial f(x,t)}{\partial t} \Rightarrow \text{The change of particles over time}$$

The right hand side represents

$$D\frac{\partial^2 f(x,t)}{\partial x^2} \Rightarrow D \text{ times the change in particles over the change in area}$$

D (the mass diffusivity) is

$$D \Rightarrow \text{Area divided by time}$$

Now let's look back at our function $\lambda f(\lambda x, \lambda^2 t)$. Let's say we want to make $\lambda^2 t$ = 1 by picking a λ to make that statement true. If we chose $\lambda = \frac{1}{\sqrt{t}}$ we'll end up with

$$\frac{1}{\sqrt{t}} f\left(\frac{x}{\sqrt{t}}, 1\right)$$

Know the workaround this to get it all into a function of 1 variable (which is the ultimate goal) is to look at really what $\frac{x}{\sqrt{t}}$ means. If we take the square root of D we get length divided by square root time. So to get it all into one variable we set λ to be $\lambda = \frac{1}{\sqrt{Dt}}$. That gives us

$$\frac{1}{\sqrt{Dt}} f\left(\frac{x}{\sqrt{Dt}}, \frac{1}{D}\right)$$

Then if we set $z = \frac{x}{\sqrt{Dt}}$ we get

$$\lambda f(\lambda x, \lambda^2 t) = \frac{1}{\sqrt{Dt}} \bar{f}(z) \tag{2}$$

Now we need to take derivatives to set up the diffusion equation. Start by taking the derivative with respect to time. Let's take the inner derivative that we get from the chain rule

$$\frac{\partial z}{\partial t} = -\frac{x}{2\sqrt{D}} \cdot t^{-\frac{3}{2}}$$

Now for the whole derivative

$$\begin{aligned}
\frac{\partial \lambda f(\lambda x, \lambda^2 t)}{\partial t} &= \frac{\partial \bar{f}(z)}{\partial z} \cdot \frac{\partial z}{\partial t} \\
&= \frac{1}{\sqrt{Dt}} \cdot \frac{\partial \bar{f}(z)}{\partial t} + \bar{f}(z) \cdot \frac{\partial}{\partial t}\left(\frac{1}{\sqrt{Dt}}\right) \\
&= -\frac{1}{\sqrt{Dt}} \cdot \frac{\partial \bar{f}(z)}{\partial z} \cdot \frac{x}{2\sqrt{D}} \cdot t^{-\frac{3}{2}} - \bar{f}(z) \cdot \frac{1}{2\sqrt{(D)}} \cdot t^{-\frac{3}{2}} \\
&= -\frac{t^{-\frac{3}{2}}}{2\sqrt{D}}\left(z \cdot \frac{\partial \bar{f}(z)}{\partial z} + \bar{f}(z)\right)
\end{aligned}$$

Now let's take the derivative with respect to x. Let's start by finding that inner derivative first

$$\frac{\partial z}{\partial x} = \frac{1}{\sqrt{Dt}}$$

Now for the whole derivative

$$
\begin{aligned}
\frac{\partial \lambda f(\lambda x, \lambda^2 t)}{\partial t} &= \frac{1}{\sqrt{Dt}} \cdot \frac{\partial \bar{f}(z)}{\partial z} \cdot \frac{1}{\sqrt{Dt}} \\
&= \frac{1}{Dt} \cdot \frac{\partial \bar{f}(z)}{\partial z}
\end{aligned}
$$

Now take the second derivative with respect to x

$$
\frac{\partial^2 \lambda f(\lambda x, \lambda^2 t)}{\partial x^2} = \frac{1}{Dt} \cdot \frac{\partial^2 \bar{f}(z)}{\partial z^2} \cdot \frac{1}{\sqrt{Dt}}
$$

Our new diffusion equation becomes

$$
-\frac{t^{-\frac{3}{2}}}{2\sqrt{D}} \left(z \cdot \frac{\partial \bar{f}(z)}{\partial z} + \bar{f}(z) \right) = \frac{\partial^2 \bar{f}(z)}{\partial z^2} \cdot \frac{1}{\sqrt{D}} t^{-\frac{3}{2}}
$$

Then we can cancel out D and t terms

$$
-\frac{z}{2} \cdot \frac{\partial \bar{f}(z)}{\partial z} - \frac{1}{2} \bar{f}(z) = \frac{\partial^2 \bar{f}(z)}{\partial z^2}
$$

Then we can rewrite it is in this form

$$
\frac{d^2 \bar{f}(z)}{dz^2} + \frac{1}{2} z \cdot \frac{d\bar{f}(z)}{dz} + \frac{1}{2} \bar{f}(z) = 0
$$

We can pack z and $\bar{f}(z)$ by using the multiplicative rule of derivatives in reverse

$$
\frac{d^2 \bar{f}(z)}{dz^2} + \frac{1}{2} \frac{d}{dz} (z \bar{f}(z)) = 0
$$

Now integrate to get

$$
\int \frac{d^2 \bar{f}(z)}{dz^2} + \frac{1}{2} \frac{d}{dz} (z \bar{f}(z)) dz \Rightarrow \frac{d\bar{f}(z)}{dz} + \frac{1}{2} z \bar{f}(z) = c
$$

We are going to use a special trick. We know that $\bar{f}(z)$ is an even function, that is because it is equally likely to move in either direction which means that

$$
\left. \frac{d\bar{f}(z)}{dz} \right|_{z=0} = 0
$$

That means that $c = 0$ and our new equation becomes

$$
\frac{d\bar{f}(z)}{\bar{f}(z)} + \frac{1}{2} z dz = 0
$$

Then let's isolate the first fraction

$$
\frac{d\bar{f}(z)}{\bar{f}(z)} = -\frac{1}{2} z dz
$$

Notice that it is differential of the log of the function $\bar{f}(z)$

$$d \ln \bar{f}(z) = -\frac{1}{2} z dz$$

Then integrate to get

$$\int d \ln \bar{f}(z) dz \Rightarrow \ln \bar{f}(z) - \ln \bar{f}(0) = -\frac{z^2}{4}$$

Then exponentiate to get

$$\bar{f}(z) = \bar{f}(0) e^{-\frac{z^2}{4}} \tag{3}$$

Now we have the number of particles at z. Also keep in mind that we set

$$z = \frac{x}{\sqrt{Dt}}, \quad D = \frac{1}{2\tau} \int_{-\infty}^{\infty} \Delta^2 \phi(\Delta) d\Delta$$

This says that equation 2 is really the strength of the particles moving.

Now we need to find $\bar{f}(0)$. We do this by applying in our initial conditions. If we consider that we have m particles and it is large enough for the statistics and probability to work out.

$$\bar{f}(0) \int_{\mathbb{R}} e^{-\frac{z^2}{4} dz} = m$$

Now we are going to use a trick, we want to use a u-substitution that will give us a normal density. If we set this for our substitution

$$z = \sqrt{2} u, \quad dz = \sqrt{2} du$$

Then we get

$$\sqrt{2} \bar{f}(0) \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du = m$$

This look like a normal but it is missing the $\sqrt{2\pi}$ so we know that integral must give us

$$\sqrt{2} \bar{f}(0) \sqrt{2\pi} = m$$

Now solve for $\bar{f}(0)$

$$\bar{f}(0) = \frac{m}{\sqrt{4\pi}}$$

Now plug that into eq.3

$$\bar{f}(z) = \frac{m}{\sqrt{4\pi}} e^{-\frac{z^2}{4}}$$

Now we plug that back into eq.2

$$\lambda f(\lambda x, \lambda^2 t) = \frac{1}{\sqrt{Dt}} f\left(\frac{x}{\sqrt{Dt}}\right)$$

Now expand the function to get

$$\lambda f(\lambda x, \lambda^2 t) = \frac{m}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

This is giving us the number of particles at location x after t amount of time passes.

Now what we are going to do is use this functions at different x locations. In this case we will let x vary and call this function $\psi(x) = f(x, 0)$. Now we can aggregate all of the locations as

$$f(x,t) = \int_{-\infty}^{\infty} \psi(z) \frac{1}{\sqrt{4\pi Dt}} \cdot e^{-\frac{(x-z)^2}{4Dt}} dz$$

We have now found the solution. We give names to these functions. Everything in exponent is Green's Function and the $\psi(z)$ is the impulse function.

3.2 Analytical Solution

Now we will find the analytical solution to the diffusion equation. Let's start off with

$$\frac{\partial f(x,t)}{\partial t} = D \cdot \frac{\partial^2 f(x,t)}{\partial x^2}$$

Let's set theses initial conditions

$$f(x, 0) = f(x), \quad \forall x \in [0, L]$$

$$f(0, t) = f(L, t) = 0, \quad \forall t > 0$$

Then we are going to use the separation of variables which assumes $f(x, t) = X(x)T(t)$. And let's plug that into the diffusion equation

$$\frac{1}{D} \cdot \frac{T'(t)}{T(t)} = \frac{X''(x)}{X(x)}$$

Then we can make these systems of equations

$$X''(x) + \lambda X(x) = 0 \tag{4}$$

$$T'(t) + D\lambda T(t) = 0 \tag{5}$$

When we solve the linear ODE and taking into consideration the initial conditions

$$X(x) = C_1 e^{\sqrt{-\lambda x}} + C_2 e^{-\sqrt{-\lambda x}}$$

Now we are going to break it up into how λ changes. For when $\lambda < 0$ we get $C_1 = C_2 = 0$.

Now for the $\lambda = 0$ we get

$$X(x) = C_1 x + C_2$$

Then for the $\lambda > 0$ case we get

$$X(x) = C_1 \cos(\sqrt{\lambda}x) + C_2 \sin(\sqrt{\lambda}x)$$

When we plug in the boundary condition we get

$$X(0) = C_1 = 0$$

$$X(L) = C_2 \sin(\sqrt{\lambda}L) = 0$$

When we solve λ we get

$$\lambda_n = \left(\frac{\pi n}{L}\right)^2$$

Plug will give us a

$$X(x) = C_n \sin\left(\frac{\pi n}{L}x\right)$$

Now working out eq.5 we end up with

$$T'(t) + D\left(\frac{\pi n}{L}\right)T(t) = 0$$

That gives us

$$T(t) = B_n e^{-D\left(\frac{\pi n}{L}\right)^2 t}$$

Now lets put it all together

$$f(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{\pi n}{L}x\right) e^{-D\left(\frac{\pi n}{L}\right)^2 t}$$

Now we need to solve for A_n which becomes

$$A_n = \frac{2}{L}\int_0^L f(\delta)\sin\left(\frac{\pi n}{L}\delta\right) d\delta$$

Then all that becomes

$$f(x,t) = \sum_{n=1}^{\infty}\left(\frac{2}{L}\int_0^L \sin\left(\frac{\pi n}{L}\delta\right) d\delta\right)\sin\left(\frac{\pi n}{L}x\right) e^{-D\left(\frac{\pi n}{L}\right)^2 t}$$

4 Diffusion Convection Equation

This is more of an extension of the diffusion equation where we add in a convection / drift. Let's start with our diffusion equation

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2}$$

The diffusion convection equation becomes

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2} - \mu\frac{\partial f(x,t)}{\partial x}$$

We can think of this as a force or a flow, so they move randomly but they have a preferred direction. Normally when we thought of the diffusion equation back in section 1. We were finding the taylor series for $f(x+\Delta,t)$. But for this case $f(x+\Delta,t) \neq f(x-\Delta,t)$ that is because there is now a preference.

We can rewrite our diffusion equation to be

$$f(x,t+\tau)dx = dx\int_{-\infty}^{\infty} f(x+\Delta,t)\phi(-\Delta)d\Delta$$

The $\phi(-\Delta)$ happens because of the drift. Then we can get rid of the differentials.

$$f(x,t+\tau) = \int_{-\infty}^{\infty} f(x+\Delta,t)\phi(-\Delta)d\Delta \tag{1}$$

Now we can rewrite each side of the equation as

$$f(x,t+\tau) = f(x,t) + \frac{\partial f(x,t)}{\partial t}\tau$$

$$f(x+\Delta,t) = f(x,t) + \frac{\partial f(x,t)}{\partial x}\Delta + \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x^2}\Delta^2$$

We can plug those into equation

$$f(x,t) + \frac{\partial f(x,t)}{\partial t}\tau = \int_{-\infty}^{\infty}\left(f(x,t) + \frac{\partial f(x,t)}{\partial x}\Delta + \frac{1}{2}\cdot\frac{\partial^2 f(x,t)}{\partial x^2}\Delta^2\right)\phi(-\Delta)d\Delta$$

Then we can break up the integral

$$f(x,t)+\frac{\partial f(x,t)}{\partial t}\tau = \int_{-\infty}^{\infty}\phi(-\Delta)d\Delta+\frac{\partial f(x,t)}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(-\Delta)d\Delta)+\frac{1}{2}\cdot\frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta$$

By rules of probability we know that

$$\int_{-\infty}^{\infty}\phi(-\Delta)d\Delta = 1$$

That gives us

$$f(x,t)+\frac{\partial f(x,t)}{\partial t}\tau = f(x,t)+\frac{1}{2}\cdot\frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta+\frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta$$

We can get rid of the $f(x,t)$ and get

$$\frac{\partial f(x,t)}{\partial t}\tau = \frac{1}{2}\cdot\frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta + \frac{\partial^2 f(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta$$

Then we can rewrite this as

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2 f(x,t)}{\partial x^2}\cdot\frac{1}{2\tau}\cdot\int_{-\infty}^{\infty}\Delta^2\phi(-\Delta)d\Delta + \frac{\partial f(x,t)}{\partial x}\cdot\frac{1}{\tau}\int_{-\infty}^{\infty}\Delta\phi(-\Delta)d\Delta$$

Now we are going to use another u-substitution except we are doing this to change what we are integrating to respect with

$$\frac{\partial^2 f(x,t)}{\partial t} = -\frac{\partial =^2 f(x,t)}{\partial x^2} \cdot \frac{1}{2\tau} \int_{\infty}^{-\infty} \Delta^2 \phi(\Delta) d\Delta + \frac{\partial f(x,t)}{\partial x} \cdot \frac{1}{\tau} \int_{\infty}^{-\infty} \Delta \phi(\Delta) d\Delta$$

Then to switch those integrating bounds we get

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2 f(x,t)}{\partial x^2} \cdot \frac{1}{2\tau} \int_{-\infty}^{\infty} \Delta^2 \phi(\Delta) d\Delta - \frac{\partial f(x,t)}{\partial x} \cdot \frac{1}{\tau} \int_{-\infty}^{\infty} \phi(\Delta) d\Delta$$

We can see that the first integral is the diffusion component

$$D = \frac{1}{2\tau} \int_{-\infty}^{\infty} \Delta^2 \phi(\Delta) d\Delta$$

We can call the second integral the drift

$$\mu = \frac{1}{\tau} \int_{-\infty}^{\infty} \Delta \phi(\Delta) d\Delta$$

Now we get the diffusion convection equation

$$\frac{\partial f(x,t)}{\partial t} = D \frac{\partial f(x,t)}{\partial x^2} - \mu \frac{\partial f(x,t)}{\partial x}$$

Lets use this approximation method

$$\begin{aligned} \frac{f(x,t+\tau) - f(x,t)}{\tau} &= D \frac{f(x+\Delta,t) - 2f(x,t) + f(x-\Delta,t)}{\Delta^2} - \mu \frac{f(x+\Delta,t) - f(x-\Delta,t)}{2\Delta} \\ &= \frac{2D}{\Delta^2} \left(\frac{f(x+\Delta,t) + f(x-\Delta,t)}{2} - f(x,t) \right) + \frac{\mu}{2\Delta} (f(x-\Delta,t) - f(x+\Delta,t)) \end{aligned}$$

The first part shows that the particles diffuse from higher concentration to lower concentration. And the second part shows that location x will gain the particles from the left hand side and lose them to the right hand side at the same rate in accordance with the drift.

5 Diffusion Convection Equation solution

We have the diffusion equation and its solution

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2}, \;\; f(x,t) = \frac{m}{\sqrt{2\pi Dt}}e^{-\frac{x^2}{4Dt}}$$

Now we want to solve the equation for the diffusion convection equation. We need to find the variable transformation, in this case we want the transformation to reflect convection

$$y = x - \mu t$$

Let's find the derivatives

$$\frac{\partial y}{\partial x} = 1, \;\; \frac{\partial y}{\partial t} = -\mu$$

Now take the derivatives of $f(x,t)$ and keep in mind that we need the chain rule

$$\begin{aligned}\frac{\partial f(x,t)}{\partial x} &= \frac{\partial f(y,t)}{\partial y}\cdot\frac{\partial y}{\partial x}\\ &= \frac{\partial f(y,t)}{\partial y}\end{aligned}$$

Then take the second derivative

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial^2 f(y,t)}{\partial y^2}$$

Now for the derivative with respect to t via total derivative

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} &= \frac{\partial f(y,t)}{\partial t} + \frac{\partial f(y,t)}{\partial t}\cdot\frac{\partial y}{\partial t}\\ &= \frac{\partial f(y,t)}{\partial t} - \mu\frac{\partial f(y,t)}{\partial y}\end{aligned}$$

Now plug that into our convection diffusion equation

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2} - \mu\frac{\partial f(x,t)}{\partial x}$$

$$\Downarrow$$

$$\frac{\partial f(y,t)}{\partial t} - \mu\frac{\partial f(y,t)}{\partial y} = D\frac{\partial^2 f(y,t)}{\partial y^2} - \mu\frac{\partial f(y,t)}{\partial y}$$

$$\Downarrow$$

$$\frac{\partial f(y,t)}{\partial t} = D\frac{\partial^2 f(y,t)}{\partial y^2}$$

Now to write out the solution we can use this fact $f(x,0) = f(y,0)$ so lets write out the solution in terms of y

$$f(y,t) = \frac{m}{\sqrt{4\pi Dt}}e^{-\frac{y^2}{4Dt}}$$

Then use the definition of y to get the final result

$$f(x,t) = \frac{m}{\sqrt{4\pi Dt}}e^{-\frac{(x-\mu t)^2}{4Dt}}$$

Appendix A: Showing that the transformation is invariant

To show that they are invariant we can apply u and v to $f(x,t)$ to make sure that they still hold. That gives us this function $f(\lambda x, \lambda^2 t)$. Now take the derivatives

$$\frac{du}{dx} = \lambda, \quad \frac{du}{dt} = \lambda^2$$

Now take partial derivatives of f and by chain rule

$$\begin{aligned}\frac{\partial f(v,u)}{\partial t} &= \frac{\partial f(v,u)}{\partial u} \cdot \frac{du}{dt} \\ &= \lambda^2 \cdot \frac{\partial f(v,u)}{\partial u}\end{aligned}$$

Now take the other partial derivative

$$\begin{aligned}\frac{\partial f(v,u)}{\partial x} &= \frac{\partial f(v,u)}{\partial v} \cdot \frac{dv}{du} \\ &= \lambda \cdot \frac{\partial f(v,u)}{\partial v}\end{aligned}$$

Now take the partial derivative with respect again for f_x to satisfy the heat equation

$$\begin{aligned}\frac{\partial^2 f(v,u)}{\partial x^2} &= \lambda \cdot \frac{\partial}{\partial x}\left(\frac{\partial f(v,u)}{\partial v}\right) \\ &= \lambda \cdot \frac{\partial^2 f(v,u)}{\partial v^2} \cdot \frac{dv}{du} \\ &= \lambda^2 \cdot \frac{\partial^2 f(v,u)}{\partial v^2}\end{aligned}$$

We are able to set up the diffusion equation under our transformation

$$\frac{\partial f(v,u)}{\partial u} = D \cdot \frac{\partial^2 f(v,u)}{\partial v^2}$$

Then if we plug in u and v we get

$$\frac{\partial f(\lambda x, \lambda^2 t)}{\partial t} = D \cdot \frac{\partial^2 f(\lambda x, \lambda^2 t)}{\partial x^2}$$

Appendix B: Finding the transformation

Let's try and find our transformation using an arbitrary one. If we scale each variable by some constants $a, b \in \mathbb{R}$ giving us

$$v = bx, \quad u = ct \Rightarrow f(v,u) = f(bx, ct)$$

Then we try working out the partial derivatives again, if we start with the t derivative

$$\begin{aligned}\frac{\partial f(v,u)}{\partial t} &= \frac{\partial f(v,u)}{\partial u} \cdot \frac{du}{dt} \\ &= \frac{\partial f(v,u)}{\partial u} \cdot c\end{aligned}$$

And for the x partial derivative which we will have to take twice

$$\begin{aligned}\frac{\partial f(v,u)}{\partial x} &= \frac{\partial f(v,u)}{\partial v} \cdot \frac{du}{dv} \\ &= \frac{\partial f(v,u)}{\partial v} \cdot b^2 \\ \frac{\partial^2 f(v,u)}{\partial x^2} &= b^2 \cdot \frac{\partial}{\partial x}\left(\frac{\partial f(v,u)}{\partial x}\right) \\ &= b^2 \cdot \frac{\partial^2 f(v,u)}{\partial x^2}\end{aligned}$$

Then if we set those equal to each other we see

$$c\frac{\partial f(v,u)}{\partial x} = b^2 \cdot \frac{\partial^2 f(v,u)}{\partial x^2}$$

Then we can see that $b^2 = c$. That is where we get the λ from the transformation.

C. Interchanging the integral (u-substitution)

We have this integral

$$\int_{\mathbb{R}} af(\lambda x, \lambda^2 t)dx$$

If we make this substitution $v = \lambda x$. Let's work it out like how we would work out a normal u-substitution

$$v = \lambda x$$

$$dv = \lambda dx$$

$$dx = \frac{dv}{\lambda}$$

Then plug it into integral

$$\int_{\mathbb{R}} af(v, \lambda^2 t)\frac{1}{\lambda}dv$$

Introductory concepts for Term Structure Modeling

Diego Alvarez
diego.alvarez@colorado.edu

May 2021

Author's note

This paper is meant to serve as a framework for future models as well as laying the groundwork for term structure models. Another reason for publishing this paper is showcase my skills regarding finance and mathematics. Please read through, and of course if any mistakes do appear feel free to contact me.

Definitions

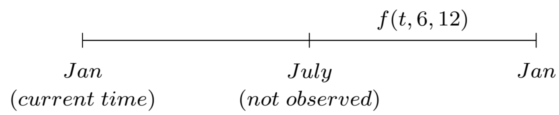
Throughout this paper there will be a series of different interest rate terms being thrown around. For ease of use and to avoid confusion the terms will be defined below.

- coupon - when a bond is issued it has coupons that are interest payments, they are also paid on a schedule.
- zero coupon bond - This is a bond that has no coupon and is bought at a discount value and when it matures it is worth \$1.
- spot rate - think of this as rate of return for without collecting coupons if you sell the bond immediately. For example if we had a \$1,000 zero coupon bond that had 2 years left until maturity and the current value is \$950 the spot rate is 2.59%. The spot rate is calculated as

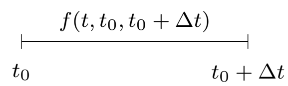
  $$\left(\frac{FV}{CV}\right)^{\frac{1}{\tau}} - 1$$

  For the most part we will commonly look at zero-coupon bonds. For those who are familiar with financial terms, the spot rate will be yield to maturity (YTM) of the zero-coupon bond. That is not the case for other coupon-bearing bonds.

- forward rate - the forward rate is an agreed upon interest rate for a bond that starts on a future date. For example if the it is January the 6-month forward rate for a maturity of 1 year is the interest rate for a 6 month bond starting in July.

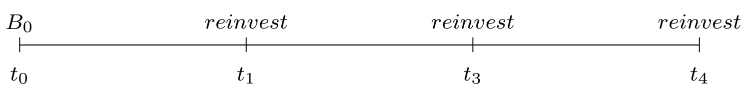


- instantaneous forward rate - an agreed upon rate for some future date date but in this case the settlement date is the next instant.



for an infinitesimally small Δt

- Bank account - This is a financial investment that we make that involves putting money into a bank account with some interest rate and then reinvesting the money over some interval.



- short rate - this refers to an interest rate that is set at some time i. Throughout this paper the short rate will be the instantaneous forward rate, this is done for ease of use with notation later on.

Notation

This will help clarify some of the functions and independent variables that we use throughout this paper

- Standard notation for functions of 2 independent variables:

  For a function g that will either model price or interest rate there are usually two independent variables associate with it. The function will usually take the form $g(t, T)$. t will be the time variable, and T will be the maturity.[1]

- Standard notation for a function of 3 independent variables:

  These functions will be used for the forward rate. It will usually be expressed as $f(t, t_i, t_{i+1})$. The first variable, t is time variable, and the t_i and t_{i+1}) are for the future date. Forward rates refer an interest rate that starts in the future. That would mean that for $f(t, t_i, t_{i+1})$ the interest rate payment starts on t_i and ends on t_{i+1} [2]

- The switch from 3 independent variables to 2 independent variables with the same function.

  Forward rates can reference a future rate. For example if we have a forward rate for the future interval (t_1, t_2) the forward rate is expressed as $f(t, t_1, t_2)$. When we are at t_1 the forward rate becomes $f(t, t_2)$.

- $y(t, T)$: the spot rate
- $P(t, T)$: this is a pricing function. It models the price of a fixed-income security with maturity T.
- $B(t, T)$: This is our bank rate. Via definition of bank-rate stated above this is the interest rate that we would receive for depositing our money.
- $f(t, T)$: This is the forward rate for some maturity T
- $r(t)$ and r_i: This is the short rate. The short rate is the continuously compounded for an infinitesimally small period

$$r(t) = f(t, t)$$

[1] I like to think of T as the variable that references the bond by maturity, other than the instantaneous forward rate its more of a variable that references the bond and doesn't change.

[2] In this case think of t_1 and t_2 as the referencing component of the security like how T was the referencing security for the function of two independent variables.

Motivations for commonly used practices

Use of zero coupon bonds

The motivation for the zero coupon bond is that we know its final value. Throughout the paper the zero coupon bond will have the value of $1.

$$P(T,T) = 1$$

Knowing the value at maturity allows us to circumvent the time aspect of zero coupon bonds. When working with other bonds the expected value at maturity is unknown. Using that feature of the zero coupon bond makes it possible to solve for other variables.

The use of instantaneous forward rates

The use of the instantaneous forward rate is so we can build continuous interest rates within the bonds. This is done because we can discount the zero coupon bond as a series of forward rates and then decrease the time interval of each forward rate. The instantaneous forward rate also helps with finding other continuous interest rates, and is necessary for setting up differential equations with interest rates. When using a $\sum$ that is in reference to the simple forward rate, and when using the $\int$ that is in reference for instantaneous rate.

Something to keep in mind is that in application there is no such thing as an instantaneous forward rate because that would imply that we lend and receive interest rate payment and reinvest faster than the speed of light. Working out this problem of instantaneous forward rates will come up in later papers, specifically the LIBOR market model paper and the HJM Framework paper.

Finding the instantaneous forward rate

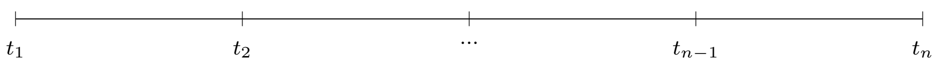
Start with the value of the zero coupon bond, knowing that the value of the bond at maturity, T is worth \$1.

$$P(T,T) = 1$$

Then express the zero coupon bond at time t through its spot value. Think of this more as representing the discounted value of the zero coupon bond.

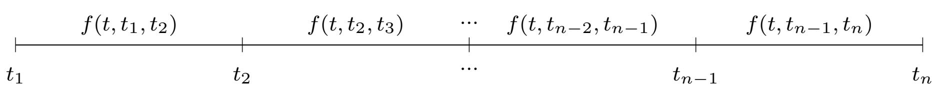
$$P(t,T) = e^{-y(t,T)(T-t)} \tag{1}$$

It is possible to translate the zero coupon bonds so that they are expressed in only forward rates. Let's break our time interval into a series of discrete interval from $t_0, ..., t_n$.



Now for each interval (t_i, t_{i+1}) make its associated forward rate. For an interval (t_i, t_{i+1}) the associated forward rate will be $f(t, t_i, t_{i+1})$

For all of the intervals the forward rates are



Then via the pricing function $P(t,T)$ the zero coupon bond will become all of the forward rates between the interval multiplied by the final price of the bond, \$1. In financial terms the zero coupon bond is being discounted to all of the forward rates of each interval. For a zero coupon bond with n subintervals, it will have $n-1$ forward rates associated with it, that is because each forward rate represents an interval rather than a timestamp. Notice that the function for the forward rates are of two variables, that is because we observe each forward rate at their respective starting date. Also because all of our time intervals are the same length our difference in time $(T-t)$ from eq. 1 becomes Δt

Our pricing function where we discount the final value of the zero coupon bond \$1 to the each forward rate becomes

$$P(t,T) = e^{-f(t,t_0)\Delta t} e^{-f(t,t_1)\Delta t} ... e^{-f(t,t_{n-1})\Delta t} \cdot 1$$

Then represent those forward rates as a summation

$$P(t,T) = e^{-\sum_{i=0}^{n-1} f(t,t_i)\Delta t}$$

Then translate those into the instantaneous forward rate by making the interval infinitesimally small and taking the limit.[3]

$$P(t,T) = e^{-\int_t^T f(t,u)du} \tag{2}$$

Now setting the two sides equal and solving for the interest rate, $y(t,T)$ we get. (See Appendix A)

$$y(t,T) = \frac{1}{T-t}\int_t^T f(t,u)du \tag{3}$$

Now the spot interest rate is represented as the average of instantaneous forward rate. [4]

[3]Almost like reversing the definition of an integral $\int f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i*)\Delta x$

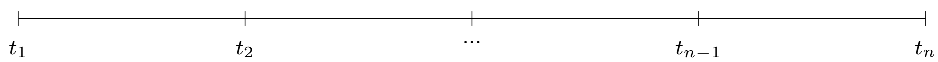
[4]In this case the referencing part is constantly changing which makes sense because the instantaneous forward rate the forward rate for an infinitesimally small time period.

Finding the instantaneous rate for the bank account

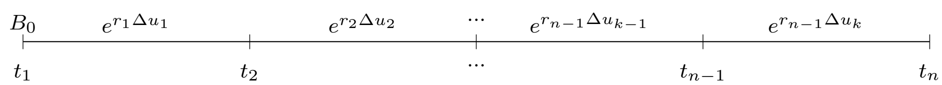
We start with \$1 at t_0 and we lend the money at an interest rate r_i for interval i. The the standard bank rate formula is.

$$B_0 e^{r_i \Delta u_i} \tag{4}$$

For the instantaneous forward rate we split the interval into n subintervals



Then we reinvest all the money over each interval and using k as the index.



We can represent the bank account over k intervals as

$$B_0 e^{r_1 \Delta u_1} e^{r_2 \Delta u_2} ... e^{r_k \Delta u_k}$$

Similar to finding the instantaneous forward rate we make a summation of interest rates

$$B_0 e^{r_1 \Delta u_1} e^{r_2 \Delta u_2} ... e^{r_k \Delta u_k} = B_0 e^{\sum_{i=0}^{k} r_i \Delta u_k}$$

Then integrate to get short rate

$$B_0 e^{\sum_{i=0}^{k} r_i \Delta u_k} = B_0 e^{\int_0^t r_u du}$$

The bank rate is[5]

$$B_t = e^{\int_0^t r_u du} \tag{5}$$

The value of the Bank account can't be expressed in terms of forward rates because the forward rate are determined ex-ante so we can't go back into time and observe the forward rates.

[5] We can drop the B_0 because it is \$1. In other words the B_0 doesn't change anything via identity property for multiplication

The zero coupon bond in terms of the short rate

To find this we will need the future value of the short rate. We use the risk-neutral valuation. Via risk neutral valuation if we take the price of the zero coupon bond at maturity and scale it by the bank account at T, the expected value under our risk neutral measure Q is the same as its current value. [6]

$$\frac{P(t,T)}{B(t)} = \mathbb{E}\left[\frac{P(T,T)}{B(T)}\middle|\mathscr{F}_t\right] \tag{6}$$

Now solve for $P(t,T)$ (see appendix B)

$$P(t,T) = \mathbb{E}\left[\frac{B(t)}{B(T)}\middle|\mathscr{F}_t\right] \tag{7}$$

We know that the bank rate at T is

$$B_T = e^{\int_0^T r_u du}$$

The bank rate at t is

$$B_t = e^{\int_0^t r_u du}$$

Knowing that $t < T$ (see appendix C)

$$\frac{B_T}{B_t} = e^{\int_t^T r_u du}$$

We can invert the equation above to get

$$\frac{B_t}{B_T} = e^{-\int_t^T r_u du}$$

Plugging that into the eq.7

$$P(t,T) = \mathbb{E}\left[e^{-\int_t^T r_u du}\right] \tag{8}$$

That gives us the value of the zero coupon bond in terms of the short rate.

[6] Using this risk neutral valuation is **very important** for the HJM model. It is also is important because it allows us to combine 2 equations together and start solving for parts of each.

Finding the instantaneous forward for the zero coupon bond

The price of bond from eq.2 is

$$P(t,T) = e^{-\int_t^T f(t,u)du}$$

Now solve for the $f(t,T)$ to get (see appendix D)

$$\frac{-d}{dT} \ln P(t,T) = f(t,T) \tag{9}$$

Eq. 9 is the instantaneous forward rate for the zero coupon bond. We can also think of it as the limit of the simple forward rate. It can also represent the present value of zero coupon bond discounted by the forward rate as some $T + \Delta$

$$P(t,T,T+\Delta) = P(t,T)e^{-f(t,T,T+\Delta)\Delta} \tag{10}$$

Then when we isolate $f(t,T,T+\Delta)$ (see appendix E)

$$f(t,T,T+\Delta) = -\frac{\ln(P(t,T+\Delta) - \ln P(t,T)}{\Delta} \tag{11}$$

Then to find the instantaneous forward rate, (in this case it would look like $f(t,T,T)$ if we represented it in terms of 3 variables). We take the limit as Δ goes to 0.

$$\lim_{\Delta \to 0} f(t,T,T+\Delta) = -\lim_{\Delta \to 0} \frac{\ln P(t,T+\Delta) - \ln P(t,T)}{\Delta}$$

The equation above is the definition of a derivative which becomes

$$f(t,T,T) = f(t,T) = -\frac{d}{dT} \ln P(t,T)$$

Comparing differentials of short rate and instantaneous forward rate

Let's call $r(t)$ our short rate and by definition it is also our instantaneous forward rate.

$$r(t) = \lim_{T \to t} f(t,T) = f(t,t) \tag{12}$$

Then we can find the differential of that (see appendix F)

$$dr(t) = df(t,T)\bigg|_{T=t} + \frac{\partial f(t,T)}{\partial T}\bigg|_{T=t} dt \tag{13}$$

This will be later used for two HJM model and then for short rate models.

Appendix A: Solving for $y(t,T)$ with instantaneous rate with zero coupon bond

The value of the zero coupon bond from eq.1 is

$$P(t,T) = e^{-y(t,T)(T-t)}$$

And the value of the zero coupon bond in terms of instantaneous forward rate (which is eq.2)

$$P(t,T) = e^{-\int_t^T f(t,u)du}$$

Now set them equal

$$e^{-y(t,T)(T-t)} = e^{-\int_t^T f(t,u)du}$$

That becomes

$$y(t,T)(T-t) = \int_t^T f(t,u)du$$

Then solve for $y(t,T)$

$$y(t,T) = \frac{1}{T-t}\int_t^T f(t,u)du$$

Appendix B: Solving for $P(t,T)$ under the risk neutral measure

We start with $P(t,T)$ scaled by $B(t)$ under the risk neutral measure

$$\frac{P(t,T)}{B(t)} = \mathbb{E}^Q\left[\frac{P(T,T)}{B(T)}\middle|\mathscr{F}_t\right]$$

Knowing that the value of $P(T,T) = 1$

$$\frac{P(t,T)}{B(t)} = \mathbb{E}^Q\left[\frac{1}{B(T)}\middle|\mathscr{F}_t\right]$$

And then solve for $P(t,T)$ by multiplying out $B(t)$ and passing it through the expected value

$$P(t,T) = \mathbb{E}^Q\left[\frac{B(t)}{B(T)}\middle|\mathscr{F}_t\right]$$

Appendix C: Getting the proportion of the bank rate

Knowing that the bank rate is

$$B_t = e^{\int_0^t r_u du}$$

We can express B_t and B_T individually as

$$B_t = e^{\int_0^t r_u du} \text{ and } B_T = e^{\int_0^T r_u du}$$

We also assume that $t < T$ so when we set up this fraction we get

$$\frac{B_T}{B_t} = \frac{e^{\int_0^T r_u du}}{e^{\int_0^t r_u du}}$$

We can break up the upper integral

$$\int_0^T r_u du = \int_0^t r_u du + \int_t^T r_u du$$

Plugging that into the original fraction

$$\frac{B_T}{B_t} = \frac{e^{\left(\int_0^t r_u du + \int_t^T r_u du\right)}}{e^{\int_0^t r_u du}}$$

Then via exponent rules we get

$$\frac{B_T}{B_t} = \frac{e^{\int_0^t r_u du} \cdot e^{\int_t^T r_u du}}{e^{\int_0^t r_u du}}$$

Now reduce the fraction

$$\frac{B_T}{B_t} = e^{\int_t^T r_u du}$$

Appendix D: Solving for the $f(t,T)$ for the zero coupon bond

Start with the price of the zero coupon bond that we found in eq.2

$$P(t,T) = e^{-\int_t^T f(t,u)du}$$

Take the log of both sides and move the negative sign over

$$-\ln P(t,T) = \int_t^T f(t,u)du$$

Then take a T-derivative

$$\frac{-d}{dT}\ln P(t,T) = \frac{d}{dT}\int_t^T f(t,u)du \tag{14}$$

Then via Leibniz Integral rule Rule which is (for some arbitrary function $f(x,z)$)

$$\frac{\partial}{\partial z}\int_a^b f(x,z)dx = \int_a^b \frac{\partial}{\partial z}f(x,z)dx + f(b,z)\frac{\partial b}{\partial z} - f(a,z)\frac{\partial a}{\partial z}$$

Applying Leibniz Integral rule to our integral (eq.14) we get

$$\frac{d}{dT}\int_t^T f(t,u)du = \int_t^T \frac{\partial}{\partial T}f(t,u)du + f(t,T)\frac{\partial}{\partial T}T - f(T,T)\frac{\partial}{\partial T}t \tag{15}$$

The partial derivative on right goes to 0.

$$\frac{d}{dT}\int_t^T f(t,u)du = \int_t^T \frac{\partial}{\partial T}f(t,u)du + f(t,T)\frac{\partial}{\partial T}T$$

Let's look at the integral on the right hand side. The derivative of a definite integral is 0.

$$\int_t^T \frac{\partial}{\partial T}f(t,u)du = 0$$

Working out the partial derivative on the right hand side we get

$$\frac{\partial}{\partial T}T = 1$$

That makes eq.15 become

$$\frac{d}{dT}\int_t^T f(t,u)du = f(t,T)$$

Appendix E: finding the forward rate of the zero coupon bond for $T + \Delta$

from eq. 9 we have

$$P(t, T, T + \Delta) = P(t, T)e^{-f(t,T,T+\Delta)\Delta}$$

divide out

$$e^{f(t,T,T+\Delta)\Delta} = \frac{P(t, T)}{P(t, T + \Delta)}$$

Then take the log

$$f(t, T, T + \Delta)\Delta = \ln\left(\frac{P(t, T)}{P(t, T + \Delta)}\right)$$

Then divide out the Δ and use log rules

$$f(t, T, T + \Delta) = -\frac{\ln P(t, T, T + \Delta) - \ln P(t, T)}{\Delta}$$

Appendix F: Finding the differential of the short rate

We first start with the short rate, which is the continuous forward rate. Starting with eq. 12

$$r(t) = \lim_{T \to t} f(t, T) = f(t, t)$$

When we want to find the differential of the short rate we are increasing the instantaneous forward rate for some time Δt. [7]

$$f(t + \Delta t, t + \Delta t)$$

That means that when we take find the differential for $r(t)$ we need to take derivative with respect to two independent variables of the forward rate above with respect to T and evaluate at t

$$dr(t) = df(t, T)\Big|_{T=t} + \frac{\partial f(t, T)}{\partial T}\Big|_{T=t} dt$$

[7] In financial terms if we plot all of the forward rates for the timeline it will look like a staircase going from left to right top to bottom. The bank rate follows the leftmost diagonal (top to bottom), so to work that out in terms of forward rates we need to move down and to right to get to the next "step". That is why we move in 2 directions for Δt

Heath-Jarrow-Morton Framework

Diego Alvarez

September 2021

Contents

1 Introduction

The Heath-Jarrow-Morton (HJM) Model models a dynamics of an instantaneous forward. We focus on developing the model for a generic maturity T. We also assume no arbitrage. In this case a cornerstone to this model is finding the volatility term. The volatility term determines the type of stochastic process.

2 Motivations

Motivation for Risk Neutral Measure

This model is built upon a stochastic differential equation of the form

$$df(t,T) = \mu(t,T)dt + \sigma_f(t,T)dW_t$$

Later on we will find other expressions for the parameters in the SDE. That will be done by building them out through the term structure tools discussed in the older paper. Then from there we can bring the two sub models together under the risk neutral pricing.

The risk neutral measure that we found is

$$\frac{P(t,T)}{B(t)} = \mathbb{E}^Q \left[\frac{P(T,T)}{B(T)} \middle| \mathscr{F}_t \right]$$

From here if we knew of the volatility of an security's process then we can write the dynamics out by using the risk neutral measure. The risk neutral measure allows us to draw assumptions about how the security will perform at maturity. Under the risk neutral measure the expected return is the risk-free rate.

3 Initial Model

We first assume that there is no arbitrage

Start with a general stochastic differential equation of this form

$$df(t,T) = \mu(t,T)dt + \sigma_f(t,T)dW_t \tag{1}$$

The expected return under the risk free measure will be the instantaneous forward rate, but the instantaneous forward rate is not a tradable security. Instead we'll need the dynamics of the zero coupon of the same maturity. The SDE for the zero coupon bond is. (See appendix A for full calculation)

$$dP(t,T) = r_t P(t,T)dt + \sigma_P(t,T)P(t,T)dW_t \tag{2}$$

Recall that to *transfer* from the instantaneous forward rate to the zero coupon bond we get

$$f(t,T) = -\frac{d}{dT} \ln P(t,T)$$

If we assume that we can switch the derivative and differential and make this differential

$$df(t,T) = -\frac{d}{dT} d \ln P(t,T) \tag{3}$$

Then by Ito's lemma we get (See Appendix B for full calculation)

$$d \ln P(t,T) = \frac{1}{P(t,T)} dP(t,T) - \frac{1}{2} \cdot \frac{1}{P(t,T)^2} dP(t,T)^2 \tag{4}$$

We can find those fractions on the right hand side manipulating eq.2. Let's first start by dividing by $P(t,T)$.

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sigma_P(t,T)dW_t \tag{5}$$

Now to find the fraction in the second term we take the square. Using the fact that $dt^2 = dt$ and $dW_t^2 = 0$, (results from Ito's lemma) we get (See appendix C for full calculation)

$$\frac{dP(t,T)^2}{P(t,T)} = \sigma_P(t,T)^2 dt \tag{6}$$

Now putting those fractions that we found from eq.5 and eq.6 we get

$$d \ln P(t,T) = r_t dt + \sigma_P(t,T) dW_t - \frac{1}{2} \cdot \sigma_P(t,T)^2 dt \tag{7}$$

Now take a derivative with respect to T

$$\frac{d}{dT} d \ln P(t,T) = \frac{d}{dT} \sigma_P(t,T) dW_t - \sigma_P(t,T) \frac{d}{dT} \sigma_P(t,T) dt$$

We can use the fact that the differential of the forward is negative of what we just found (eq.3) to get this

$$df(t,T) = -\frac{d}{dT} d \ln P(t,T) = \sigma_P(t,T) \frac{d}{dT} \sigma_P(t,T) dt - \frac{d}{dT} \sigma_P(t,T) dW_t$$

We know how the dynamics of the forward under the risk neutral measure

$$df(t,T) = \sigma_P(t,T) \frac{d}{dT} \sigma_P(t,T) dt - \frac{d}{dT} \sigma_P(t,T) dW_t \tag{8}$$

Now we can relate eq.8 to eq.1. We can do this because both of their differentials are in regards to the same variables. Therefore we can make this claim

$$\sigma_f = -\frac{d}{dT} \sigma_P(t,T), \quad \mu(t,T) = \sigma_P(t,T) \frac{d}{dT} \sigma_P(t,T)$$

We can rewrite the σ_P by undoing the derivative with via integration

$$\int_t^T \sigma_t(t,u) du + C = -\sigma_P(t,T)$$

We can use the fact that the volatility at maturity is 0 $\sigma_P(t,T) = 0$ therefore $C = 0$ leaving us with

$$\sigma_P(t,T) = -\int_0^T \sigma_f(t,u) du \tag{9}$$

Now working out the $\mu(t,T)$ for eq.1 and eq.8. We can then plug in the value of σ_P into our $\mu(t,T)$

$$\mu(t,T) = \sigma_P(t,T) \int_t^T \sigma_f(t,u) du$$

Now we have the dynamics of the instantaneous forward under the risk neutral measure

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t$$

Really what we can extract is that we needed to work out the volatility, without that we wouldn't have been able to translate eq.1 to eq.8 or work out. We were able to make our assumptions about the volatility because of the measure that we had set before we started making the model.

4 Under the T-Forward measure

Let's start by defining this new measure. This method is also called the change of Numeriare. It is used for pricing techniques that involve random discount factors using forward measures. We start off with this processing which is a martingale under the risk neutral.

$$\frac{V(0)}{B(0)} = \mathbb{E}^Q\left[\frac{V(t)}{B(t)}\right] \tag{1}$$

Instead of using the bank account we will use the price of zero coupon

$$\frac{V(0)}{P(0,T)} = \mathbb{E}^T\left[\frac{V(t)}{P(t,T)}\right] \tag{2}$$

Now we will find new ways to rewrite eq.1 and eq.2. We can do this by finding $V(0)$ for each

$$V(0) = \mathbb{E}\left[\frac{B(0)}{B(t)}V(t)\right], \quad V(0) = \mathbb{E}^T\left[\frac{P(0,T)}{P(T,T)}V(t)\right]$$

We can express then use the definition of expected value and keep in mind that that each equation is under their respective probability measure

$$V(0) = \int \frac{B(0)}{B(t)}V(t)dQ, \quad V(0) = \int \frac{P(0,T)}{P(t,T)}V(t)dP$$

They are both talking about the price of the asset therefore we can make this claim

$$\frac{B(0)}{B(t)}dQ = \frac{P(0,T)}{P(t,T)}dP^T$$

Now we can rewrite this as the derivative of the new probability measure with respect to the old one.

$$\frac{dP^T}{dQ} = \frac{P(t,T)}{P(0,T)} \cdot \frac{B(0)}{B(t)} \tag{3}$$

Now we can use the differential of the log price that we got using Ito's lemma in the last section. We know that the equation is (Section 1 eq.7)

$$d\ln P(t,T) = r_t dt + \sigma_P(t,T)dW_t - \frac{1}{2}\cdot\sigma_P(t,T)^2 dt$$

Then to undo the derivative we can integrate

$$\ln P(t,T) - \ln P(0,T) = \int_0^t \left(r_u - \frac{1}{2} \cdot \sigma_P(u,T)^2 \right) du + \int_0^t \sigma_P(u,T) dW_u$$

Now we are trying to solve for $\frac{P(0,T)}{P(t,T)}$ so we can get that by using logarithm rules and then undoing those logs to get

$$\frac{P(t,T)}{P(0,T)} = e^{\int_0^t \left(r_u - \frac{1}{2} \cdot \sigma_P(u,T)^2 \right) du + \int_0^t \sigma_P(u,T) dW_u}$$

The ratio that the bank account has is

$$B(t) = B(0) e^{\int_0^t r_u du} \Rightarrow \frac{B(0)}{B(T)} = e^{-\int_0^t r_u du}$$

Now we can put those ratios into eq.3

$$\frac{dP^T}{dQ} = \frac{P(t,T)}{P(0,T)} \cdot \frac{B(0)}{B(t)} = e^{-\frac{1}{2}\int_0^t \sigma_P(u,T)^2 du + \int_0^t \sigma_P(u,T) dW_u}$$

Then using the Radon-Nikodym derivative (see appendix D) we can work backwards to get

$$dW_t^T = dW_t - \sigma_P(t,T) dt$$

Now we know σ_P (from section 1 eq.9).

$$dW_t^T = dW_t + \int_t^T \sigma_f(t,u) du dt \tag{4}$$

Now we can put that into the dynamics of the instantaneous forward under our new measure. Let's start with the one we initially found (Section 1. eq.9)

$$df(t,T) = \left(\sigma_f(t,T) \int_t^T \sigma_f(t,u) du \right) + \sigma_f(t,T) dW_t$$

Then put in the differential that we found in eq.4 and rearrange it to get dW_t

$$df(t,T) = \left(\sigma_f(t,T) \int_t^T \sigma_f(t,u) du \right) dt + \sigma_f(t,T) \left(dW_t^T - \int_t^T \sigma_f(t,u) du dt \right)$$

Then the two integrals cancel out giving us this

$$df(t,T) = \sigma_f(t,T) dW_t^T$$

This means that the instantaneous forward is a martingale under the forward measure, this is done by modeling the instantaneous forward with a zero coupon bond of the same maturity.

5 Under the $T_f > T$-Forward Measure

Usually when implementing this model we want to model a bunch of forwards $T_1, T_2, ..., T_f$. And each of those would take we would have to solve a $df(t,T), ..., df(t,T_f)$.

Instead we will take an arbitrary $T_f > T$, and then under the risk neutral measure we can write out

$$V(0) = \mathbb{E}^Q\left[\frac{B(0)}{B(t)}X_t\right], \quad V(0) = \mathbb{E}^{T_f}\left[\frac{P(0,T_f)}{P(t,T_f)}\right]$$

Then finding the Radon-Nikodym derivative becomes

$$\frac{B(0)}{B(t)}dQ = \frac{P(0,T_f)}{P(t,T_f)}dP^{T_f}, \quad \frac{dP^{T_f}}{dQ} = \frac{P(t,T_f)}{P(0,T_f)} \cdot \frac{B(0)}{B(t)}$$

The Radon-Nikodym lets us sort of change lens between the risk neutral measure for T and the risk neutral measure for T_f. Let's look at the Brownian under the T_f We have to take into account that our bounds for time is longer than T therefore we make those adjustments in the integral

$$\begin{aligned} dW_t^{T_f} &= dW_t - \sigma_P(t,T_f) \\ &= dW_t + \int_t^{T_f} \sigma_f(t,u)dudt \end{aligned}$$

Now we can add that dW_t into the our $df(t,T)$ equation

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right) + \sigma_f(t,T)\left(dW_t^{T_f} - \int_t^{T_f} \sigma_f(t,u)dudt\right)$$

We are able to cancel out the integrals because they are integrating over the same function $\sigma_f(t,u)$ with respect to the same variable u and $[t,T] \in [t,T_f]$ leaving us with

$$df(t,T) = -\sigma_f(t,u)\left(\int_t^{T_f} \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t^{T_f}$$

6 Markovian Volatility

Let's start by getting the instantaneous rate $df(t,T)$

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t$$

Now let's take the integral to undo $df(t,T)$

$$f(t,T) - f(0,T) = \int_0^t \left(\sigma_f(s,T)\int_s^T \sigma_f(s,u)du\right) + \int_0^t \sigma_f(s,T)dW_s$$

Let's assume that volatility is a deterministic function of only time and maturity $\sigma_f(t,T)$. When we look at the Brownian increments they are normally distributed (See appendix E). Now we want to make the volatility process Markovian.

To do this we need to make sure that distribution at the next interval only depends on the current distribution. This makes modelling a lot easier because we only need to look at the current distribution. Let's isolate the stochastic term

$$D(t) = \int_0^t \sigma_f(s,t)dW_s, \quad D(T) = \int_0^T \sigma_f(s,T)dW_s$$

To do this we need to check

$$D(T) - D(t) = \int_0^T \sigma_f(s,T)dW_s - \int_0^t \sigma_f(s,t)dW_s$$

We can regroup the functions by rewriting the integral but changing their bounds

$$D(T) - D(t) = \int_t^T \sigma_f(s,T)dW_s + \int_0^t (\sigma_f(s,T) - \sigma_f(s,t))dW_s \tag{1}$$

We used this trick because the first integral is a deterministic function with respect to the brownian

$$\int_t^T \sigma_f(s,T)dW_s = 0$$

Now to make the process Markov we have to make σ_f a separable function. Let's say $\sigma_f(s,T) = g(s)h(T)$. Then the eq.1 becomes

$$D(T) - D(t) = \int_t^T h(T)g(s)dW_s + \int_0^t g(s)h(T) - g(s)h(t)dW_s \tag{2}$$

$$= \int_t^T g(s)dW_s + (h(T) - h(t)) \int_0^t g(s)dW_s \tag{3}$$

Let's take a step back and recall that when we make the change for $f(s,t) = g(s)h(t)$ that it would also change $D(t)$

$$D(t) = \int_0^t h(t)g(s)dW_s$$

Which we can rewrite as

$$\frac{D(t)}{h(t)} = \int_0^t g(s)dW_s$$

Now plug that into eq.3

$$D(T) - D(t) = h(t)\int_t^T g(s)dW_s + \frac{h(T) - h(t)}{h(t)}D(t)$$

The process is now Markov. If we were to go from $t \to T$ we would only need the information of the process at time t.

7 Under the lognormal distribution

Let's start with standard HJM under the risk neutral measure

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t \tag{1}$$

If we wanted to make the volatility a lognormal process then we would rewrite it as $\sigma_f(t,T) = \sigma f(t)$. Let's plug that into eq.1

$$\begin{aligned} df(t) &= \left(\sigma f(t)\int_t^T \sigma f(t)du\right)dt + \sigma f(t)dW_t \\ &= \sigma^2 f(t)^2 \int_t^T dudt + \sigma f(t)dW_t \\ &= \sigma^2 f(t)^2 (T-t)dt + \sigma f(t)dW_t \end{aligned}$$

If we look at the deterministic portion $df(t) = \sigma^2 f(t)^2 (T-t)dt$. We can solve that ODE via separation of variables

$$\int_0^t \frac{df(u)}{f(u)^2} = \sigma^2 \int_0^t (T-u)du$$

Then work out the integral

$$-\frac{1}{f(t)} + \frac{1}{f(0)} = -\frac{\sigma^2}{2}(T-t)^2 + \frac{\sigma^2}{2}T^2$$

Then solve for $f(t)$

$$f(t) = \frac{f(0)}{1 - \sigma^2 t(T - \frac{t}{2})f(0)}$$

If we look specifically at the denominator we notice that there is a possibility that it can go to zero. As that happens the instantaneous forward $f(t)$ will go to infinity. If that was the case the price of the bond would be \$0 which which breaks the no-arbitrage constraint that we put on.

8 Discrete Setting

Let's start with the HJM dynamics under the risk neutral measure

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t$$

The HJM replicates the instantaneous forward for a fixed maturity. Think of the instantaneous forwards as this matrix

$$\begin{bmatrix} f(0,0) & f(0,1) & f(0,2) & f(0,3) & \dots \\ & f(1,1) & f(1,2) & f(1,3) & \dots \\ & & f(2,2) & f(2,3) & \dots \\ & & & \ddots & \dots \end{bmatrix}$$

At time t_0 we are modeling all of our instantaneous forwards $f(0,0), f(0,1), ..., f(0,n)$. When we go to t_1 all of our instantaneous forwards get updated to $f(1,1), f(1,2), ..., f(1,n)$. You can think of the matrix is moving down diagonally and each row is the instantaneous forwards that we use for the HJM.

Let's start by enumerating some of the forward rates, in this case we will use h for the intervals.

$$\begin{aligned}
t_1 &: B(t_1) = e^{f(0,0)h} \\
t_2 &: B(t_2) = e^{f(0,0)h} \cdot e^{f(1,1)h} \\
t_3 &: B(t_3) = e^{f(0,0)h} \cdot e^{f(1,1)h} \cdot e^{f(2,2)h} \\
&\vdots \\
t_n &: B(t_n) = e^{f(0,0)h} \cdot e^{f(1,1)h} \cdot \ldots \cdot e^{f(n-1,n-1)h}
\end{aligned}$$

We can write that out as a summation

$$B(t_i) = e^{\sum_{j=0}^{i} f(j,j)h} \tag{1}$$

Let's model a zero coupon bond that expires at t_4

$$P(t_0, t_4) = e^{-f(0,0)h - f(0,1)h - f(0,2)h - f(0,3)h} = e^{-\sum_{j=0}^{3} f(0,j)h}$$

With eq.1 we can write that as

$$\begin{aligned}
P(t_1, t_4) &= e^{-\sum_{j=1}^{3} f(1,j)h} \\
P(t_2, t_4) &= e^{-\sum_{j=2}^{3} f(2,j)h}
\end{aligned}$$

Now we can make a generic version of the equations above

$$P(t_i, T) = e^{-\sum_{j-1}^{n-1}} f(i,j)h$$

Now we can define the HJM under the risk neutral measure. We can use this identity again that we found under the risk neutral measure

$$\frac{P(t,T)}{B(t)} = \mathbb{E}^Q \left[\frac{P(S,T)}{B(S)} \middle| \mathscr{F}_t \right]$$

The discrete version of the the martingale under the risk neutral measure becomes

$$\frac{P(t_i,T)}{B(t_i)} = \mathbb{E}^Q \left[\frac{P(t_{i+1},T}{B(t_{i+1}} \middle| \mathscr{F}_i \right]$$

Now let's multiply each side by $\frac{B(t_i)}{P(t_i,T)}$ to get

$$1 = \mathbb{E}^Q \left[\frac{P(t_{i+1},T)}{P(t_i,T)} \cdot \frac{B(t_i)}{B(t_{i+1})} \middle| \mathscr{F}_i \right] \tag{2}$$

We already found $P(t_i,T)$ and $B(t_i)$. Now we need to find $P(t_{i+1},T)$ and $B(t_{i+1})$ which we can do by slightly changing the summations that we get

$$B(t_{i+j}) = e^{\sum_{j=0}^{i} f(j,j)h}, \;\; P(t_{i+1},T) = e^{-\sum_{j=i+1}^{n-1}} f(i+1,j)h$$

Let's first look at $\frac{B(t_i)}{B(t_{i+1})}$. We can rewrite $B(t_{i+1})$ as

$$B(t_{i+1}) = e^{\sum_{j=0}^{i-1} f(j,j)h + f(i,i)h}$$

We've essentially separated the $B(t_i)$ from $B(t_{i+1})$ and the fraction reduces to

$$\frac{B(t_i)}{B(t_{i+1})} = e^{-f(i,i)h}$$

Now we need to solve for zero coupon using the same method. In this case we will get

$$P(t_i,T) = e^{-f(i,i)h - \sum_{j=i+1}^{n-1} f(i,j)h}$$
$$P(t_{i+1},T) = e^{-\sum_{j=i+1}^{n-1} f(i+1,j)h}$$

When we put them into the fraction we get

$$\frac{P(t_{i+1},T)}{P(t_i,T)} = e^{-\sum_{j=i+1}^{n-1}(f(i+1,j)-f(i,j))h + f(i,i)h}$$

Then don't cancel out because the forward curves are different. When we plug that into eq.2

$$1 = \mathbb{E}^Q\left[e^{-\sum_{j=i+1}^{n-1}(f(i+1,j)-f(i,j))h + f(i,i)h} \cdot e^{-f(i,i)h} \middle| \mathscr{F}_i \right] \tag{3}$$
$$= \mathbb{E}^Q\left[e^{-\sum_{j=i+1}^{n-1}(f(i+1,j)-f(i,j))h} \middle| \mathscr{F}_i \right] \tag{4}$$

If we look inside the summation we see that it is the sum of the increments. This is really a discretized of the differential $\Delta f(t,T)$. That differential is equal to a drift term and volatility term. Our $f(i+i,j) - f(i,j)$ would be

$$f(i+1,j) - f(i,j) = \mu_{i,j}h + \sigma_{i,j}\Delta W_{i+1}$$

Then we can plug all that into eq.4

$$1 = \mathbb{E}^Q\left[e^{-\sum_{j=i+1}^{n-1}(\mu_{i,j}h + \sigma_{i,j}\Delta W_{i+1})h} \middle| \mathscr{F}_i \right] \tag{5}$$
$$= \mathbb{E}^Q\left[e^{-\sum_{j=i+1}^{n-1}\mu_{i,j}h^2 - \sum_{j=i+1}^{n-1}\sigma_{i,j}h\Delta W_{i+1}} \middle| \mathscr{F}_i \right] \tag{6}$$

Now find the mean of the Random variable in the exponent

$$\mathbb{E}\left[-\sum_{j=i+1}^{n-1} \mu_{i,j}h^2 - \sum_{j=i+1}^{n-1} \sigma_{i,j}h\Delta W_{i+1} \right] = -\sum_{j=i+1}^{n-1} \mu_{i,j}h^2$$

For variance we have

$$\begin{aligned}\operatorname{Var}\left[-\sum_{j=i+1}^{n-1}\mu_{i,j}h^2-\sum_{j=i+1}^{n-1}\sigma_{i,j}h\Delta W_{i+1}\right] &= \left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2 \operatorname{Var}[\Delta W_{i+1}] \\ &= \left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2 h\end{aligned}$$

We can then plug all that into eq.4 and we are going to use this identity, where X is a random variables

$$\mathbb{E}[e^X]=e^{\mathbb{E}[X]+\frac{1}{2}\operatorname{Var}[X]}$$

Which makes eq.4 become

$$1=e^{-\sum_{j=i+1}^{n-1}\mu_{i,j}h^2}\cdot e^{\frac{1}{2}\left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2}$$

For the equation to be true we need to the exponents to sum to 0. Therefore we can make this statement

$$\begin{aligned}\sum_{j=i+1}^{n-1}\mu_{i,j}h^2 &= \frac{1}{2}\left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2 h \\ &= \sum_{j=i+1}^{n-1}\mu_{i,j}=\frac{1}{2h}\left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2\end{aligned}$$

We want to find out the increments so first shift the upper bound

$$\sum_{j=i+1}^{n-1}\mu_{i,j}=\frac{1}{2h}\left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2$$

$$\Downarrow$$

$$\sum_{j=i+1}^{n}\mu_{i,j}=\frac{1}{2h}\left(\sum_{j=i+1}^{n}\sigma_{i,j}h\right)^2$$

Then write out the increments

$$\mu_{i,n}=\frac{1}{2h}\left(\left(\sum_{j=i+1}^{n}\sigma_{i,j}h\right)^2-\left(\sum_{j=i+1}^{n-1}\sigma_{i,j}h\right)^2\right)$$

That becomes the drift of the instantaneous forward for t_n for the interval $[i,i+1]$. If we compare it

A. Refresher on zero coupon bond price

The value of the zero coupon bond in terms of the instantaneous forward rate is

$$P(t,T) = e^{-\int_t^T f(t,u)du}$$

B. finding $d\ln P(t,T)$ and refresher on Ito's lemma

We are using this differential which relates the instantaneous forward rate to the price of the zero coupon bond

$$df(t,T) = -\frac{d}{dT} d\ln P(t,T)$$

The *problem* at hand is the $d\ln P(t,T)$. To solve this we will have to use Ito's lemma. Recall that Ito's lemma says

$$df(X) = f_X dX + \frac{1}{2} f_{XX} dX^2$$

Then applying that to $\ln X$ and knowing that

$$\frac{d}{dX}\ln X = \frac{1}{X}$$

$$\frac{d^2}{dX^2}\ln X = \frac{1}{X^2}$$

We get

$$d\ln X = \frac{1}{X}dX - \frac{1}{2}\cdot\frac{1}{X^2}dX$$

Then plug in $P(t,T)$ for X

$$d\ln P(t,T) = \frac{1}{P(t,T)}dX - \frac{1}{2}\cdot\frac{1}{P(t,T)^2}dP(t,T)^2$$

C. Solving $\frac{dP(t,T)^2}{P(t,T)^2}$ and working with dt^2 and dW_t^2

When we take the square of

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sigma_P(t,T)dW_t$$

we get

$$\frac{dP(t,T)^2}{P(t,T)^2} = r_t^2 dt^2 + \sigma_P(t,T)^2 dW_t^2$$

But we know that $dt^2 = 0$. We get this from Ito's lemma. The equation is really found by getting the solution of this integral

$$\int_0^s ds^2$$

Then by definition of an integral

$$\int_0^s ds^2 = \lim_{n\to\infty} \sum_{k=1}^{n} \Delta t_k^2$$

Here is the workaround for solving this integral. What we will do is solve it for the mean squared convergence which implies this

$$\lim_{n\to\infty} E\left[|X_n - X|^2\right] = 0$$

D. Girsanov's theorem and Radon-Nikodym Derivative

If we have a Brownian motion W_t under Q and we make this new process $Y(t)$.

$$Y(t) = \int_0^t y_u du$$

We can adjust the process W_t for $Y(t)$ which becomes

$$W_t^T = W_t - \int_0^t y_u du$$

This process is under a probability measure P^T. That probability measure comes from the Radon-Nikdoym derivative

$$\frac{dP^T}{dQ} = e^{-\frac{1}{2}\int_-^t y_u^2 du + \int_0^t y_u dW_u}$$

In differential form the two Brownian look like

$$dW_t^T = dW_t - y_t dt$$

Ultimately what we can do is that we are going to look at what we have and then translate that back to differential form via Radon-Nikodym derivative.

E. Showing that the volatility is Gaussian when volatility is a deterministic and a function of time and maturity

Let's start with our instantaneous rate

$$df(t,T) = \left(\sigma_f(t,T)\int_t^T \sigma_f(t,u)du\right)dt + \sigma_f(t,T)dW_t$$

then take the integral

$$f(t,T) - f(0,T) = \int_0^t \left(\sigma_f(s,T) \int_s^T \sigma_f(s,u)du \right) ds + \int_0^t \sigma_f(s,T)dW_s$$

this part is a linear combination of normals

$$\int_0^t \sigma_f(s,T)dW_s$$

And this part is deterministic

$$\int_0^t \left(\sigma_f(s,T) \int_s^T \sigma_f(s,u)du \right) ds$$

Really the whole thing becomes a linear transformation of deterministic functions that are Gaussian.

Efficient Frontier

Diego Alvarez
diego.alvarez@colorado.edu

May, 14th 2021

Author's note

This paper is meant to serve as a framework for future models as well as laying the groundwork for portfolio optimization and management. Another reason for publishing this paper is showcase my skills regarding finance and mathematics. Please read through, and of course if any mistakes do appear feel free to contact me.

Motivation for the Efficient Frontier

The efficient frontier is a graphical tool used to in portfolio management. It finds an optimal allocation strategy, which makes it an optimization problem. The plot of the efficient frontier finds the relationship between risk and return with different allocations of the same securities. The original motive for the efficient frontier was to find two portfolios of interest: the Markowitz Mean-Variance portfolio, and the minimum variance portfolio. More complex portfolios involve stochastic control and dynamic programming as well as solutions found using machine learning techniques. Like most financial models they usually aren't a silver bullet, there are many limitations and some problems which will be discussed at the end of this paper.

The model

Let's first start of by setting some of the initial restraints for the efficient frontier. We assume that we are working with investing in long equities and that there is no redundancy. In this case we will assume that all of the money will be invested. Although it is common practice to hold some cash as a buffer against risk, adding cash into the scope of this portfolio works somewhat against the idea of the evolution of prices. In the event that a practitioner would want to hold cash in the portfolio they can always find their desired allocation strategy and then apply that portfolio allocation model to a percentage of their portfolio.

We assume that we have N amount of risky investments to choose from. We also assume that the evolution of these prices are stochastic. The evolution a stock's price to be some stochastic function S with respect to time. For an investment $n \in N$ the price evolution will be

$$S_n(1), S_n(2), ..., S_n(T)$$

The returns of an individual security will be

$$R_n = \frac{S_n(T) - S_n(t)}{S_n(t)}$$

Now we need to classify a set of weights which will be our allocation strategy. The goal of this model is to find the optimal weights. Each weight will be the allocation size for each investment.

$$w = (w_1, ..., w_N)^T$$

Then we add this minimization constraint, because we are only allowed to invest at most 100% of our money.

$$\sum_{n=1}^{N} w_n = 1$$

The return of the portfolio can be represented as

$$R_P = \sum_{n=1}^{N} w_n R_n$$

This makes sense, our portfolio returns is each investment's return multiplied by their respective weight and then added up.

In this case we will be finding the Markowitz mean-variance portfolio and minimum-variance portfolio by looking at expected portfolio return and variance therefore we need to find are the first two moments of the portfolio.

$$\mu_P = E[R_p] = E\left[\sum_{n=1}^{N} w_n R_n\right]$$

Using the linearity of expectation and using the definition for expected returns we get

$$E\left[\sum_{n=1}^{N} w_n R_n\right] = \sum_{n=1}^{N} w_n E[R_n] = \sum_{n=1}^{N} w_n \bar{R_n}$$

Now let's find the second moment

$$\sigma_P^2 = var(R_P)$$

But we need to notice the R_P is more a collection of a series of returns therefore applying this property

$$cov(X, X) = var(X)$$

becomes

$$var(R_P) = cov(w_1 R_1, w_2 R_2) + cov(w_1 R_1, w_N R_N) + ... + cov(w_{N-1} R_{N-1}, w_N R_N)$$

Writing that as a summation we get

$$\sum_{i=1}^{N} \sum_{j=1}^{N} cov(w_i R_i, w_j R_j) = \sum_{i=1}^{N} w_i w_j cov(R_i, R_j) = \sum_{i=1}^{N} w_i w_j \sigma_{ij}$$

Essentially we can translate those covariances σ_{ij} into a covariance matrix. If we define Ω as a clearance matrix

$$\sigma_P^2 = w^T \Omega w$$

Two asset theorem

The ultimate goal of the efficient frontier is to find an optimal allocation strategy amongst a selection of investments. From a computational standpoint we plot a series of portfolios' standard deviation against their expected return. Then from there we probe to find "interesting" (in this case Markowitz mean-variance and minimum variance) portfolios. Taking a programming approach to this problem is relatively easy, it involves simulating a series of weights randomly and then using those series with historical returns and then finding the maximum or minimum of a portfolio's statistic. From a mathematical standpoint it is a bit harder but to understand how and why the shape of efficient frontier forms. To be understand that we need to look at an investment decision between two assets.

Let ρ be the correlation coefficient between the two assets and let the weight of the first asset be α which makes the second weight $1 - \alpha$

Plugging in our weights we get a usable definition for the first two moments of the portfolio

$$\mu_P = (1 - \alpha) R_1 + \alpha R_2$$

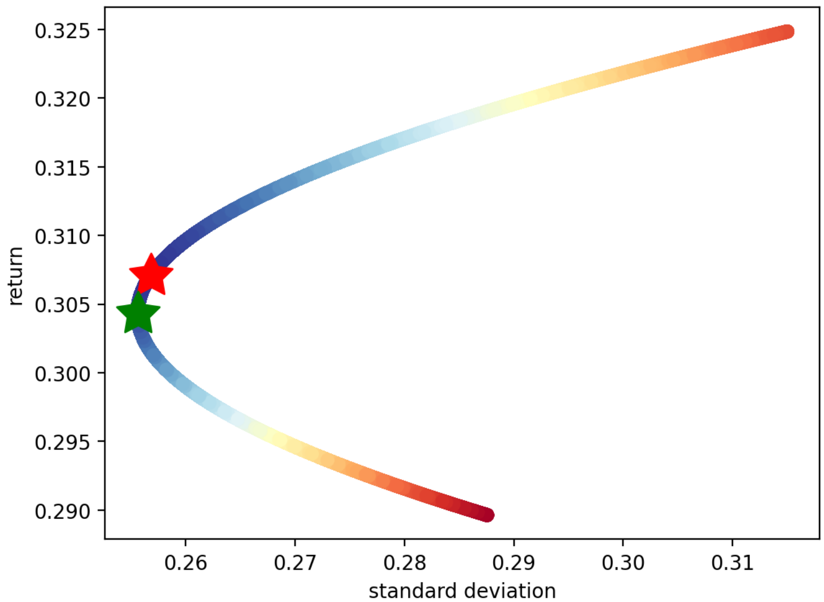
$$\sigma_P^2 = (1 - \alpha)\sigma_1^2 + 2\rho\alpha(1 - \alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2$$

$$\alpha \in [0, 1]$$

(See appendix A for second moment).

Now we should start to discuss important points that outline the shape of the efficient frontier which makes a conic shape. We can find those by solving for specific values of ρ. If we work out the scenarios of $\rho = 1$ and $\rho = -1$.

This is a sample efficient frontier that I have generated using 10 years of returns from Apple and Amazon stock using mean returns and a covariance matrix with 100,000 simulations. The prices were pulled from yaho finance and Adjusted Close was used. This was created using a web-facing GUI hosted on my GitHub written in python code. Link for app in references.



Let's work out the scenario where $\rho = 1$

$$\sigma_P(\alpha; \rho = 1) = (1 - \alpha)\sigma_1 + \alpha\sigma_2$$

(See appendix B)

Now we can work out these "edge cases" scenarios. Now we want to find the scenarios where we put all of our capital in one stock and one where we invest solely in the other. If we put the returns and standard deviation into a tuple we can work out these scenarios. From a mathematical sense we work out

$$\mu_P(\alpha = 0) = \lim_{\alpha \to 0} (1 - \alpha)R_1 + \alpha R_2 = R_1$$

$$\sigma_P(\alpha = 0, \rho = 1) = \lim_{\alpha \to 0} (1 - \alpha)\sigma_1 + \alpha\sigma_2 = \sigma_1$$

We get (σ_1, R_1) for when we let α go to 0. If we let α reach our maximum value we get

$$\mu_P(\alpha = 1) = \lim_{\alpha \to 1} (1-\alpha)R_1 + \alpha R_2 = R_2$$

$$\sigma_P(\alpha = 1, \rho = 1) \lim_{\alpha \to 1} (1-\alpha)\sigma_1 + \alpha\sigma_2 = \sigma_2$$

Which gives us (σ_2, R_2). Its obvious that we could've plugged in 0 and 1 for alpha and then connect the two points to a line using point-slope form. But the limits show that as α goes to its value it draws out a line that represents the trade off between the two securities. It is not always the case that the relationship will be linear between investments in real-world application, but for this model it is assumed.

Now let's work out the scenario of $\rho = -1$

$$\sigma_P(\alpha; \rho = 1) = |(1-\alpha)\sigma_1 - \alpha\sigma_2|$$

(See appendix C)

When we go to work out this edge case where we let α be a small number close to 0 we get

$$\lim_{\alpha \to 0} \sigma_P(\alpha; \rho = -1) = (1-\alpha)\sigma_1 - \alpha\sigma_2$$

If we set $\sigma_P = 0$ then we can solve for a specific α value

$$\alpha = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

Finding the Minimum Variance Portfolio

This portfolio is the allocation strategy with the smallest σ. If we have M portfolios to choose from the minimum variance portfolio can be found by

$$min(\sigma_1, \sigma_2, ..., \sigma_M)$$

Graphically it is the portfolio that is leftmost, and on the figure its the one with a green star. Financially it is the allocation strategy with the least amount of risk.

To mathematically find this we consider $\rho \in (0, 1)$. We are trying to find the best allocation with respect to α we need to take a partial derivative of σ_P^2 with respect to α

$$\frac{\partial \sigma_P^2}{\partial \alpha} = -2(1-\alpha)\sigma_1^2 + 2\alpha\sigma_2^2 + 2(1-2\alpha)\rho\sigma_1\sigma_2$$

Then to find the minimum variance we set the partial derivative to 0. Then solve for α.

$$\alpha = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

Finding the Markowitz Mean-Variance Portfolio

This allocation strategy is the one with the best risk-to-reward profile. I would call it “the juice is worth the squeeze” portfolio. Computationally we can find this portfolio by finding the largest Sharpe ratio. The Sharpe ratio is a portfolio statistics calculated.

$$S = \frac{\mu_P}{\sigma_P}$$

If we have M portfolios to choose from the maximum Sharpe is

$$max(S_1, S_2, ..., S_M)$$

The motivation for finding the Markowitz mean-variance portfolio is finding the best risk-to-reward portfolio. The Sharpe ratio is portfolio statistic that measures the returns vs. the risk. Therefore finding the maximized sharpe is the same as finding the Markowitz mean-variance portfolio in this model.

Financially speaking the maximized Sharpe portfolio is the best compensation for the risk that is taken on. Points on the curve further to the right is taking on more risk than what the portfolio is rewarding, and points to the left is taking on to little risk for the reward.

Mathematically we can find that by

$$min\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}\right)$$

What makes this portfolio interesting is that it is an optimization problem, which means that there is a minimizing and maximizing constraints. Other research in allocation strategies are optimization problems with different minimizing and maximizing constraints.

In the Markowitz mean-variance portfolio we have these constraints

$$\sum_{i=1}^{N} w_i R_i = \mu_P$$

$$\sum_{i=1}^{N} w_i = 1$$

To work out this optimization problem we can use the Lagrangian method for constrained optimization. We can set up the equation as

$$L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} - \lambda_1\left(\sum_{i=1}^{N} w_i - 1\right) - \lambda_2\left(\sum_{i=1}^{N} w_i R_i - \mu_P\right)$$

Then take partial derivatives with respect to w_i, λ_1, and λ_2

$$\frac{\partial L}{\partial w_i} = \sum_{j=1}^{N} \sigma_{ij} w_j - \lambda_1 - \lambda_2 R_i$$

$$\frac{\partial L}{\partial \lambda_1} = \sum_{i=1}^{N} w_i - 1$$

$$\frac{\partial L}{\partial \lambda_2} = \sum_{i=1}^{N} w_i R_i - \mu_P$$

Then we set all of those equal to 0 and start solving for the weights. When we solve the first function gives us

$$\sum_{j=1}^{N} \sigma_{ij} w_j = \lambda_1 + \lambda_2 R_i$$

We can also take note that our covariance matrix can be constructed as

$$\sum_{i=1}^{N} \sigma_{ij} = \Omega$$

Which means that we when we solve for w we get

$$w = \Omega^{-1}(\bar{1}\lambda_1 + \lambda_2 \bar{\mu})$$

$\bar{1}$ is vector filled with ones

$$\bar{1} = (1\ 1\ ...\ 1)^T$$

And $\bar{\mu}$ is the expected value of the return

$$\bar{\mu} = (R_1\ R_2\ ...\ R_N)^T$$

Now looks look at our two constraints

$$\begin{cases} \frac{\partial L}{\partial \lambda_1} = 1 \\ \frac{\partial L}{\partial \lambda_2} = \mu_P \end{cases}$$

Then we want to apply those constraints to this

$$w = \Omega^{-1}(\bar{1}\lambda_1 + \lambda_2 \bar{\mu}) = \Omega^{-1}\bar{1}\lambda_1 + \Omega^{-1}\lambda_2 \bar{\mu}$$

For the first constraint $\frac{\partial L}{\partial \lambda_1} = 1$

$$1 = \lambda_1 \bar{1}^T \Omega^{-1} \bar{1} + \lambda_2 \bar{1}^T \Omega^{-1} \bar{\mu}$$

And for the second constraint $\frac{\partial L}{\partial \lambda_2} = \mu_P$

$$\mu_P = \lambda_1 \bar{\mu}^T \Omega^{-1} \bar{1} + \lambda_2 \bar{\mu}^T \Omega^{-1} \bar{\mu}$$

We can make the notation nice with

$$\begin{cases} \eta = \bar{1}^T \Omega^{-1} \bar{1} \\ \xi = \bar{1}^T \Omega^{-1} \bar{\mu} \\ \gamma = \bar{\mu}^T \Omega^{-1} \bar{\mu} \end{cases}$$

That makes the two equations we found above become

$$1 = \lambda_1 \bar{1}^T \Omega^{-1} \bar{1} + \lambda_2 \bar{1}^T \Omega^{-1} \bar{\mu} = \eta\lambda_1 + \xi\lambda_2$$

$$\mu_P = \lambda_1 \bar{\mu}^T \Omega^{-1} \bar{1} + \lambda_2 \bar{\mu}^T \Omega^{-1} \bar{\mu} = \xi\lambda_1 + \gamma\lambda_2$$

Then we can solve for λ_1 and λ_2

$$\lambda_1 = \frac{\gamma - \xi\mu_P}{\eta\gamma - \xi^2}, \quad \lambda_2 = \frac{\eta\mu_P - \xi}{\eta\gamma - \xi^2}$$

When we have a return for the portfolio μ_P we want to find the minimum variance for those returns which gives us the best risk to reward portfolio.

$$\sigma_P^2 = w^T \Omega(\lambda_1 \Omega^{-1} \bar{1} + \lambda_2 \Omega^{-1} \bar{\mu})$$

Then working that out we get

$$\lambda_1 + \lambda_2 \mu_P = \frac{\eta\mu_P^2 - 2\xi\mu_P + \gamma}{\eta\mu_P - \xi^2}$$

Then finding the weight w becomes.

$$w = \left(\frac{\gamma - \xi\mu_P}{\eta\gamma - \xi^2} \right) \Omega^{-1} \bar{1} + \left(\frac{\eta\mu_P - \xi}{\eta\gamma - \xi^2} \right) \Omega^{-1} \bar{\mu}$$

Limitations, Problems, and Future Areas of Research

The efficient frontier was first designed to find the best risk-to-reward profile. But there are many limitations of this model. The first problem arises with using mean returns and covariance. Alternative methods work with using different risk methods such as VaR (Value-at-Risk). There is also research into covariance matrices themselves such as finding more sparse matrices for better financial descriptors and computational ease.

Other things to look into is how the efficient frontier forms with different types of stocks or asset classes. I have noticed that the efficient frontier takes a different shape if you include more securities. I hope to publish my results later on my findings.

The efficient frontier also led to the creation of the Modern Portfolio Theory which is a contested theory. MDT makes an assumption that risk and return are a trade off which isn't always the case.

As previously stated a lot of the current research involves optimization problems which can include topics of stochastic control. Also some of the optimization problems can be solved using machine learning or convex optimization.

Final Note

I hope that this paper can showcase my mathematical and financial skills when it comes to quantitative models. Although many of the areas of future research are problems that can be tackled easily with programming, I am very wary to publish my work until I have a full understanding of the mathematics. Therefore future publications on future models will be withheld until I fully understand and publish all of the mathematics behind it.

Appendix A: Second Moment

$$(w_1\ w_2)\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}\begin{pmatrix}w_1\\ w_2\end{pmatrix}$$

$$(w_1\sigma_1 + w_2\sigma_{21}\quad w_1\sigma_{12} + w_2\sigma_{22})\begin{pmatrix}w_1\\ w_2\end{pmatrix}$$

$$w_1^2\sigma_1^2$$

Appendix B: σ_P for when $\rho = 1$

$$\sigma_P^2(\alpha;\rho=1) = (1-\alpha)\sigma_1^2 + 2\alpha(1-\alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2$$

$$\sigma_P^2(\alpha;\rho=1) = ((1-\alpha)\sigma_1 + \alpha(1-\alpha)\sigma_2)^2$$

$$\sigma_P(\alpha;\rho=1) = (1-\alpha)\sigma_1 + \alpha\sigma_2$$

Appendix C: σ_P for when $\rho = -1$

$$\sigma_P^2(\alpha;\rho=1) = (1-\alpha)\sigma_1^2 - 2\alpha(1-\alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2$$

$$\sigma_P^2(\alpha;\rho=1) = ((1-\alpha)\sigma_1 - \alpha\sigma_2)^2$$

$$\sigma_P(\alpha;\rho=1) = |(1-\alpha)\sigma_1 - \alpha\sigma_2|$$

Appendix D: solving for α when $\frac{\partial\sigma_P^2}{\partial\alpha}$

$$-2(1-\alpha)\sigma_1^2 + 2\alpha\sigma_2^2 + 2(1-2\alpha)\rho\sigma_1\sigma_2 = 0$$

$$\alpha = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

References

[1] James V. Burke. "Markowitz Mean-Variance Portfolio Theory". In: *MATH 408 - Nonlinear Optimization* ().

[2] Lagrangian Methods for Constrained Optimization. "Madhavan Mukund". In: ().

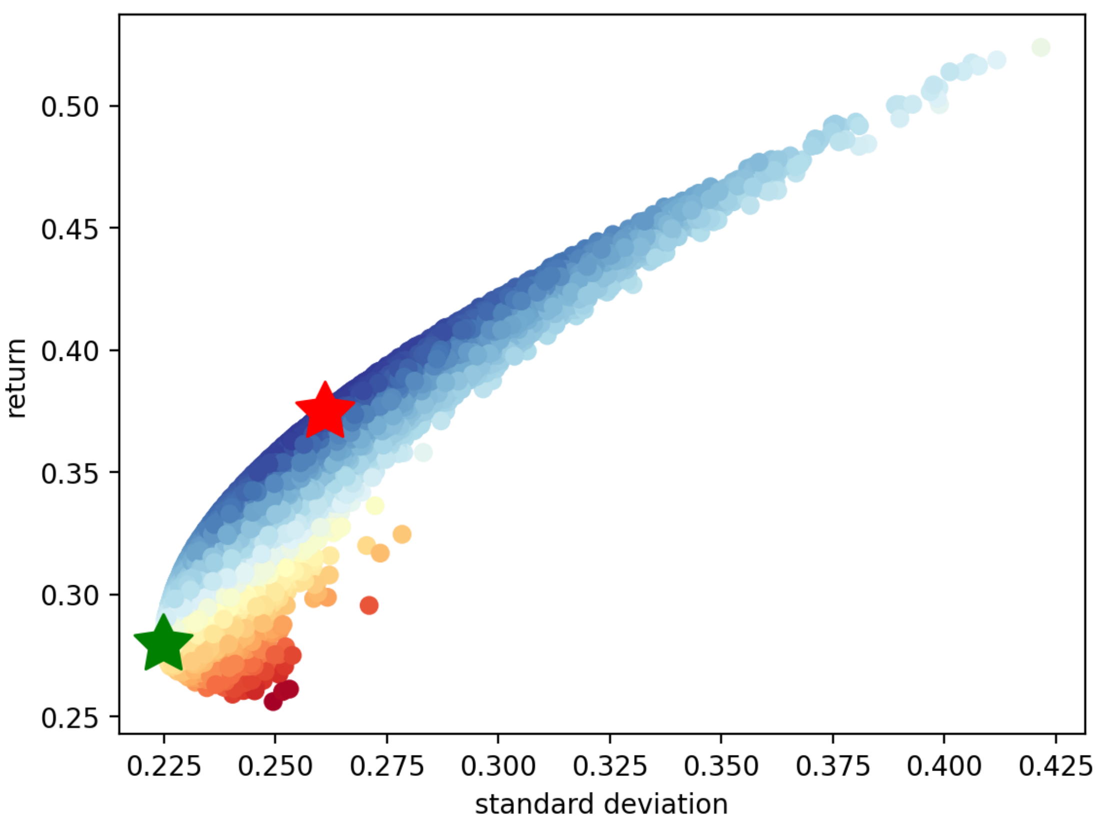
[3] Yue Kuen KWOK. "Mean-variance portfolio theory". In: *MA362* ().

[4] Markowitz Portfolio Theory. "Jingyi Zhu". In: *Introduction to Mathematical Finance II* ().

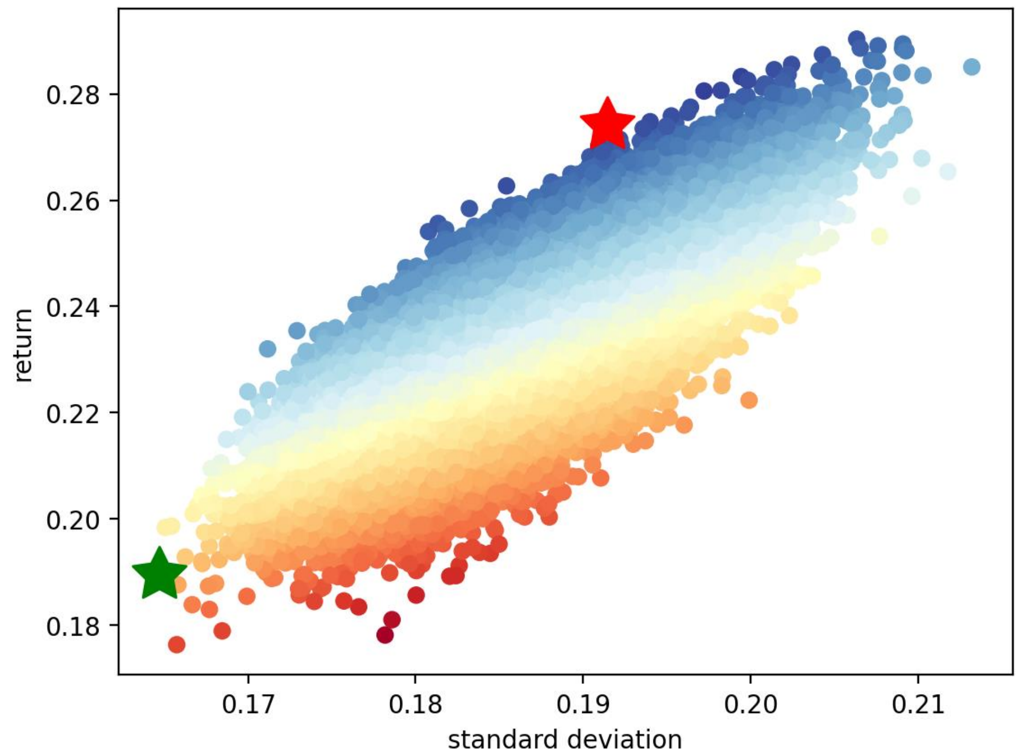
This experiment is built upon the Efficient Frontier app (link here). Although the app works as accurate testing tool the experimentation was conducted using this code base (link here). The only difference between the codebase and the app is that the codebase for the app is outfitted with the proper API to work with web-facing GUI. Both codebases use the same calculation and functions.

The efficient frontier tests use 10Y of historical adjusted close data pulled from yahoo finance. They also use the mean of daily percentage changes in adjusted close for the returns method and covariance of those returns for the risk method. Each test uses 100,000 simulations which are really 100,000 randomly generated weights and applies the mean and covariance to them to make the plots.

My initial interest in this topic was when I was trying to implement the efficient frontier to a series of stocks. Normal efficient frontiers look like this, which is an efficient frontier ran with around 7 stocks.



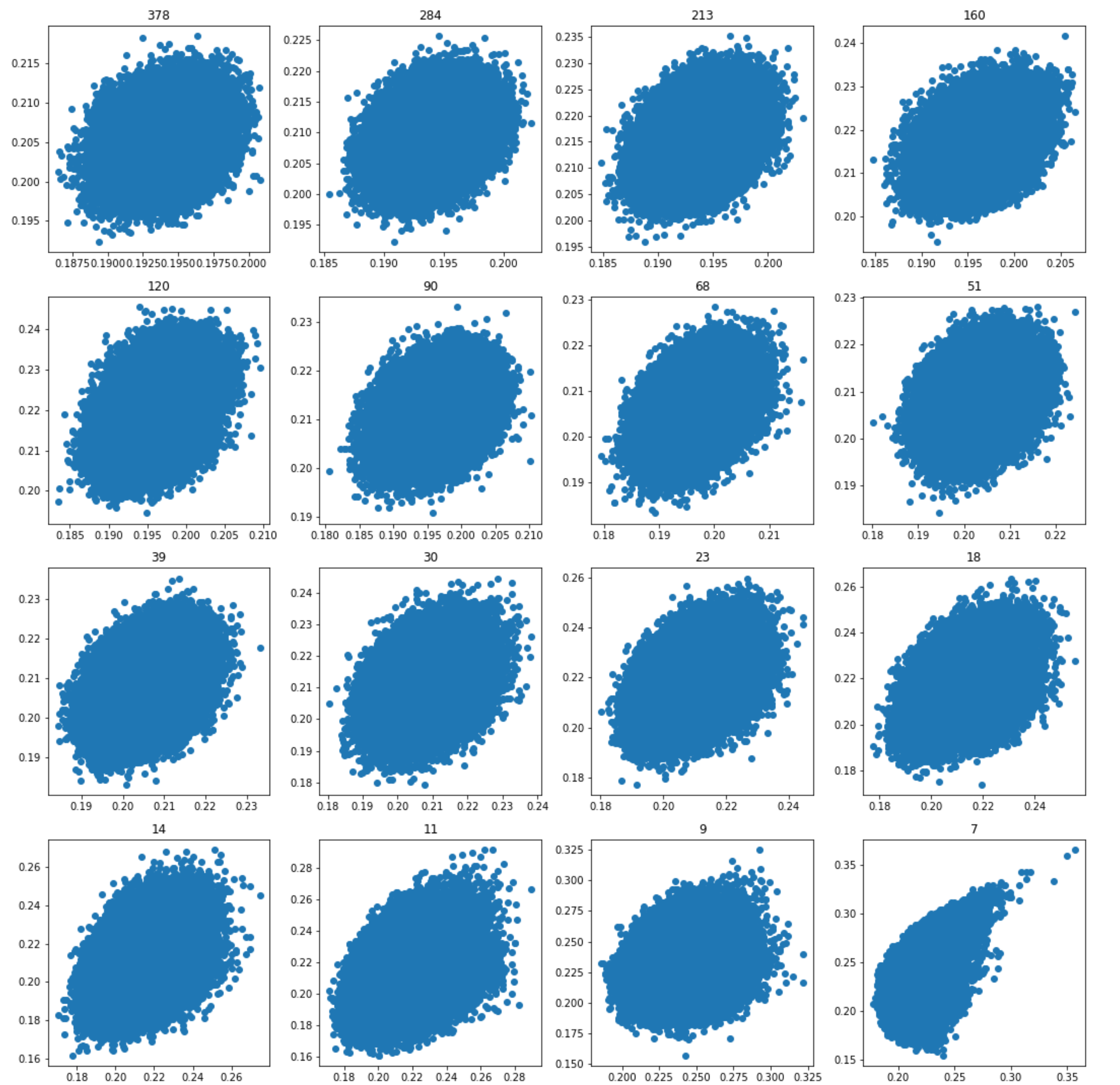
But when I tried to make an efficient frontier with around 30 stocks, I would get a graph that looks like this.



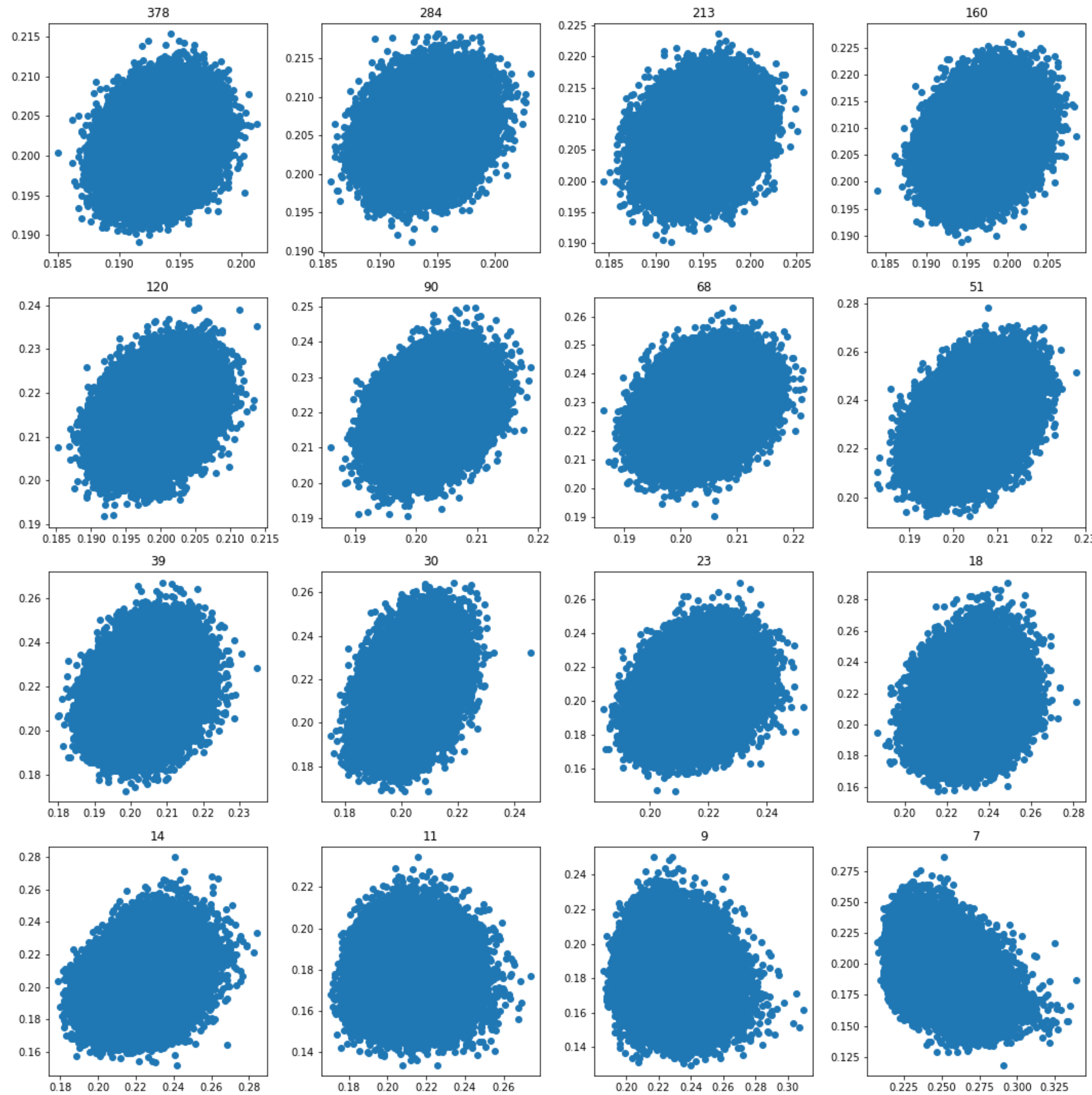
At first, I was confused and though it had something to do with the underlying securities. But after some experimentation I had a feeling, it had something to do with the number of stocks used. I started writing a program that would randomly drop stocks and then run the efficient frontier. The way that the program works is that picks several stocks and then drops a random amount (usually 25%). It then successively does that over and over, and never resamples therefore the smallest number of stocks efficient frontier has the same stocks as the largest.

I ran about 10 tests (link here). I have put 3 tests below. The title of each plot is the number of stocks used. The program also keeps track of all of the dataframes used to make the plots but for data reasons I wrote another program that keeps track of the stocks used in each test. All tests all have an associated timestamp with them.

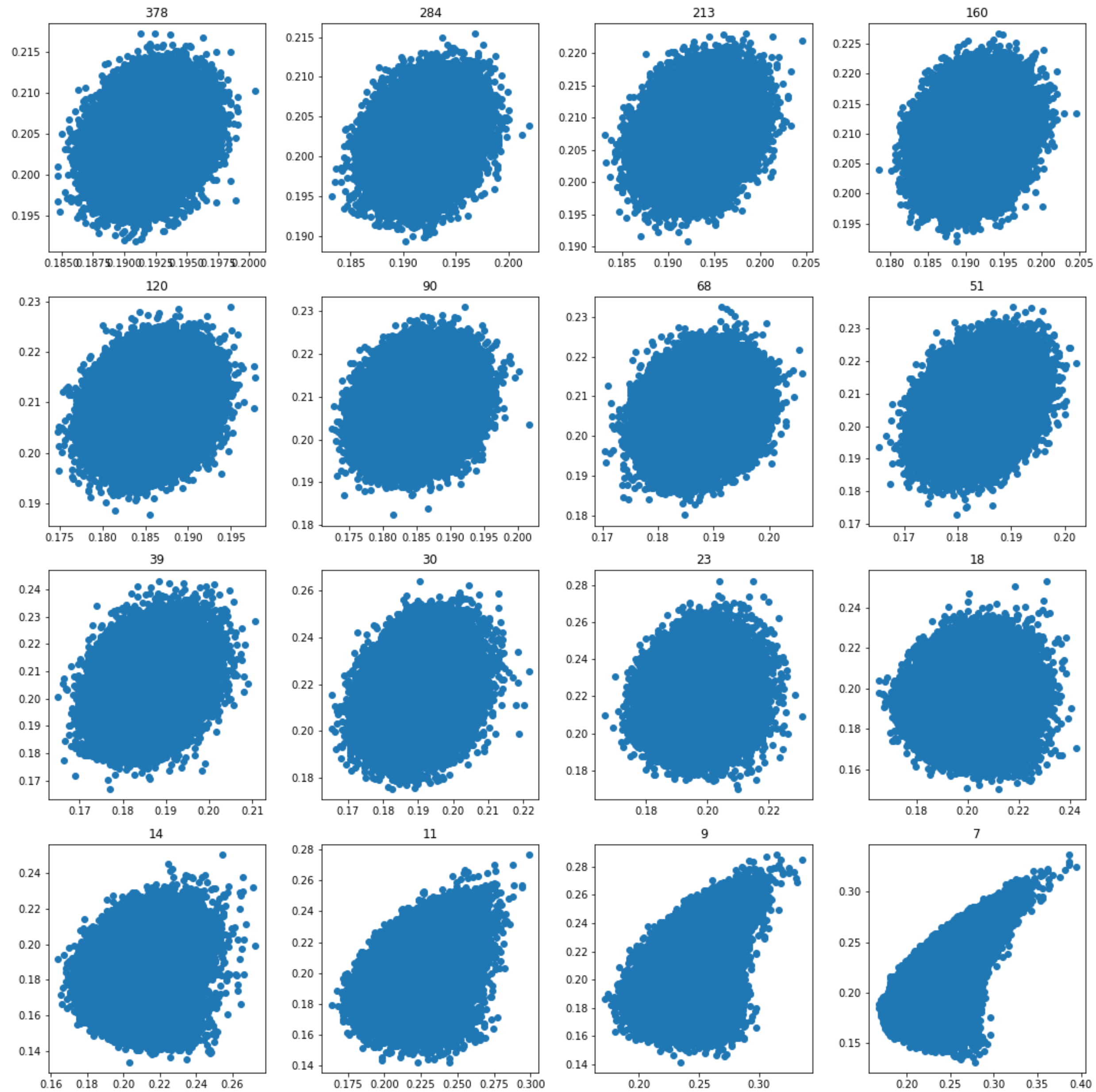
Timestamp: 1620411524.873581



Timestamp: 1620400255.83549



Timestamp: 1620376163.978867



It seems that as we drop the number of stocks used, we drop the number of outlier points and the efficient frontier starts to take shape. As of right now it seems evident that the number of stocks used determines the shape. But there are many more questions that I would like to investigate.

Future Questions & research topics:

1. Mathematical explanation for this phenomenon
2. What outlier reduction / anomaly detection methods can we use to get rid of the outliers
    a. When outliers are removed is the mean-markowitz and minimum-variance portfolio still "valid"
    b. Does the outlier reduction method work consistently across different groups of stocks and returns and risk methods?
3. Does the number of simulations play a role in getting outliers?
4. Does the proportion of outlier stay the same or does it increase / decrease as we drop the number of stocks?
5. Do different returns and risk methods yield different results?
    a. Log returns
    b. Cholesky decomposition
6. What is an acceptable level to consider a point as an outlier.

Diego Alvarez

Diego.alvarez@colorado.edu

LinkedIn | GitHub | Medium

Introduction:

I had recently built a web facing app (link here) that uses a respository I made in GitHub (link here). The repo exclusively contains python code and works with the Streamlit API. There overall goal of the app is to run time series related analysis on financial time series.

The three functions the app can perform at the time of this publication.

- Historical regime – creating a historical Markov regime switching models and then create separate time series for each regime.
- Smoothed variance probability – calculate the smoothed probability of low, medium, and high variance of returns for a financial time series.
- Continuous wavelet transform – transforming those smoothed probabilities via continuous wavelet transform to create spectrum graphs

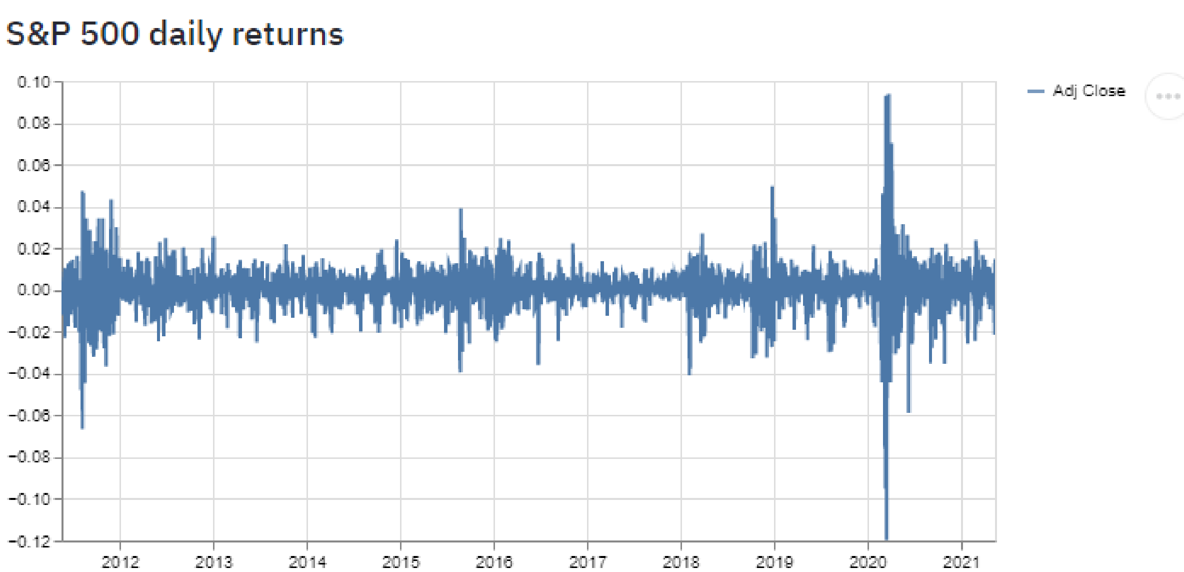
Methodology:

The data used in these experiments involves pulling data from yahoo finance by using the yfinance python API (see documentation here). The data was pulled at various time frames included (daily, weekly, monthly), and all of the prices are the Adjusted Close Price. The functions used for calculating the markov regime switching model is the python statsmodels API (see documentation here). The method used for finding the smoothed probability of the variance of returns for each regime is also built upon sample code from the statsmodels API (see 2nd example of the Kim, Nelson, and Startz (1998) Three-State Variance Switching link here). All of the code is accessible on the GitHub and all of the results are replicable using the Streamlit app.

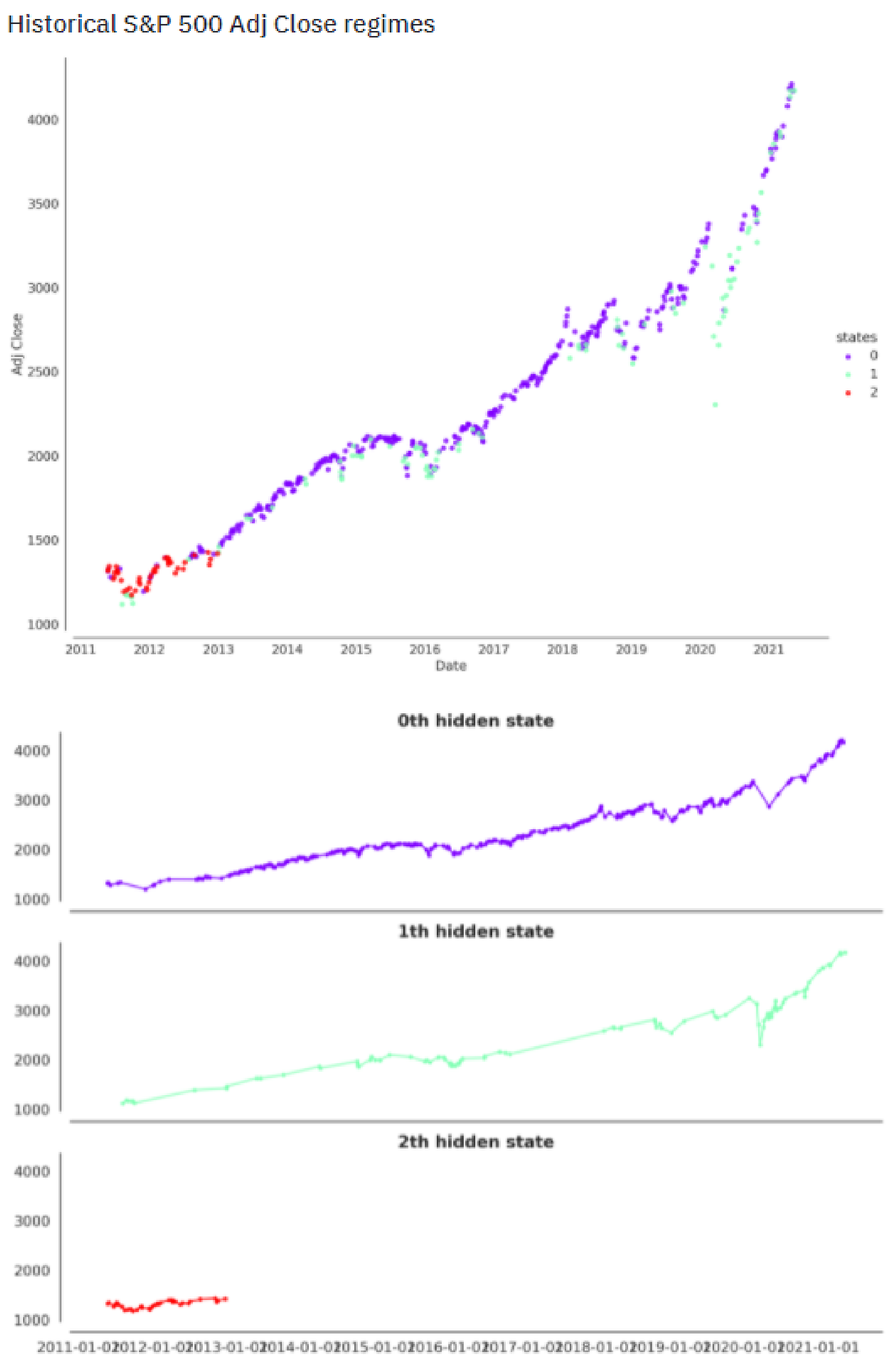
Background:

The smoothed variance probability is a Markov regime switching model. We first start with a security's adjusted close price, below is the adjusted close of the S&P 500.

Then find the percentage change. In this case the prices are the daily returns therefore we will find the daily percentage change in price, which would look something like this.

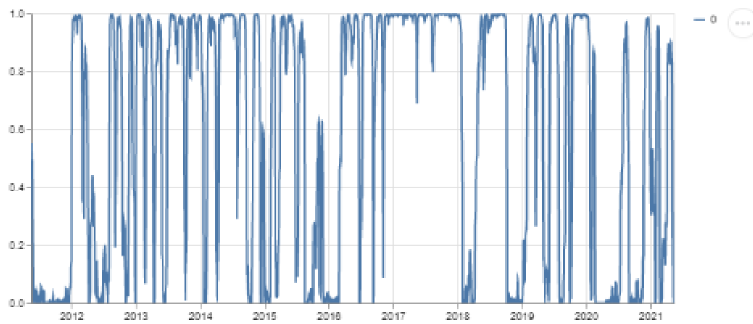


Then from there we construct a 3-regime Markov switching model dependent upon the variance of the returns. This is sort of parsing out the different timeframes of returns, and in this case, they are filtered by variance. Below is an example of a 3-regime model on price not returns, this is to give a visual idea (built in the app).

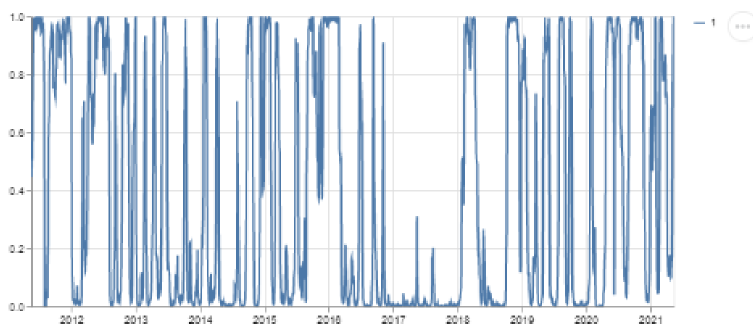


Then from there using the statsmodel method called smoothed_marginal_probabilities (see same documentation for example); that will provide three graphs that will show the probability of switching from one state to another. Finding the probability for switching regime for the daily returns of S&P 500 looks like.

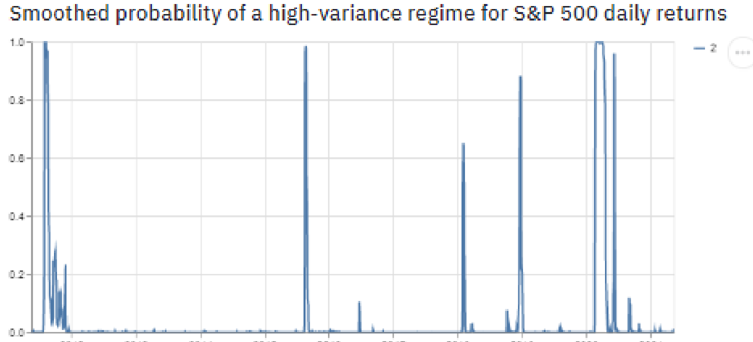
Smoothed probability of a low-variance regime for S&P 500 daily returns



Smoothed probability of a medium-variance regime for S&P 500 daily returns



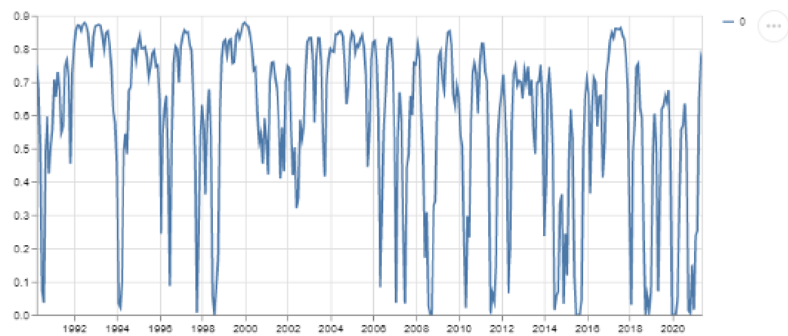
Smoothed probability of a high-variance regime for S&P 500 daily returns



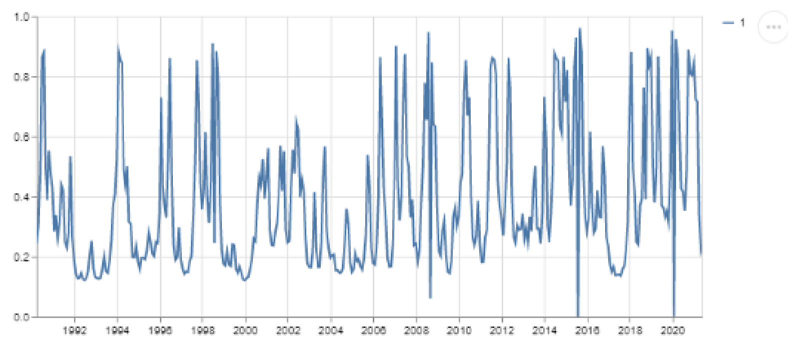
Results:

This test involves all of the historical data for CBOE VIX that exists. While searching through the smoothed marginal probability for the weekly CBOE VIX I noticed something in these graphs.

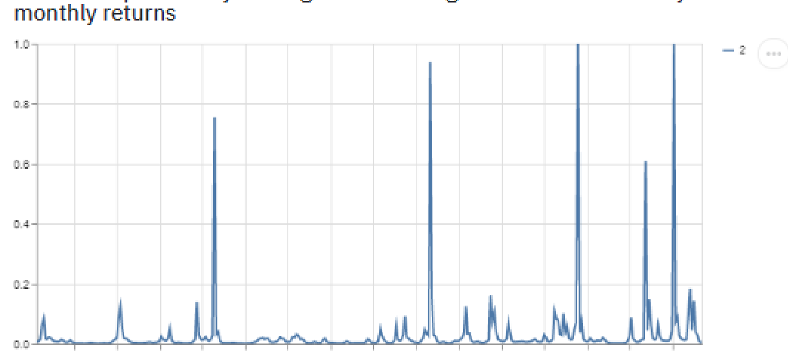
Smoothed probability of a low-variance regime for CBOE Volatility Index monthly returns



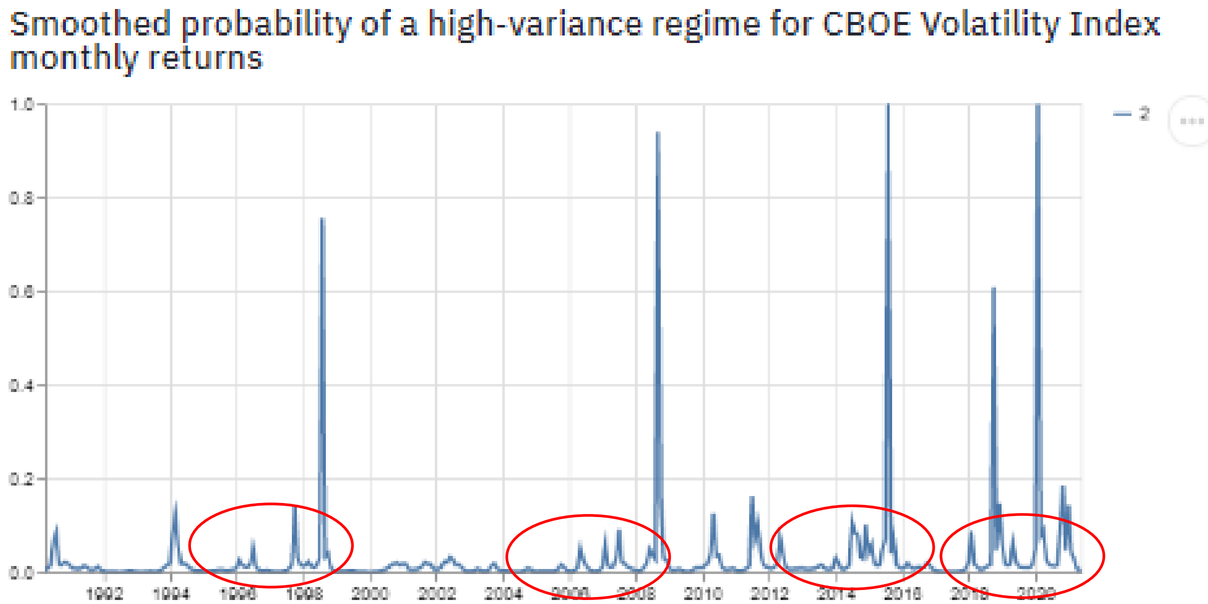
Smoothed probability of a medium-variance regime for CBOE Volatility Index monthly returns



Smoothed probability of a high-variance regime for CBOE Volatility Index monthly returns

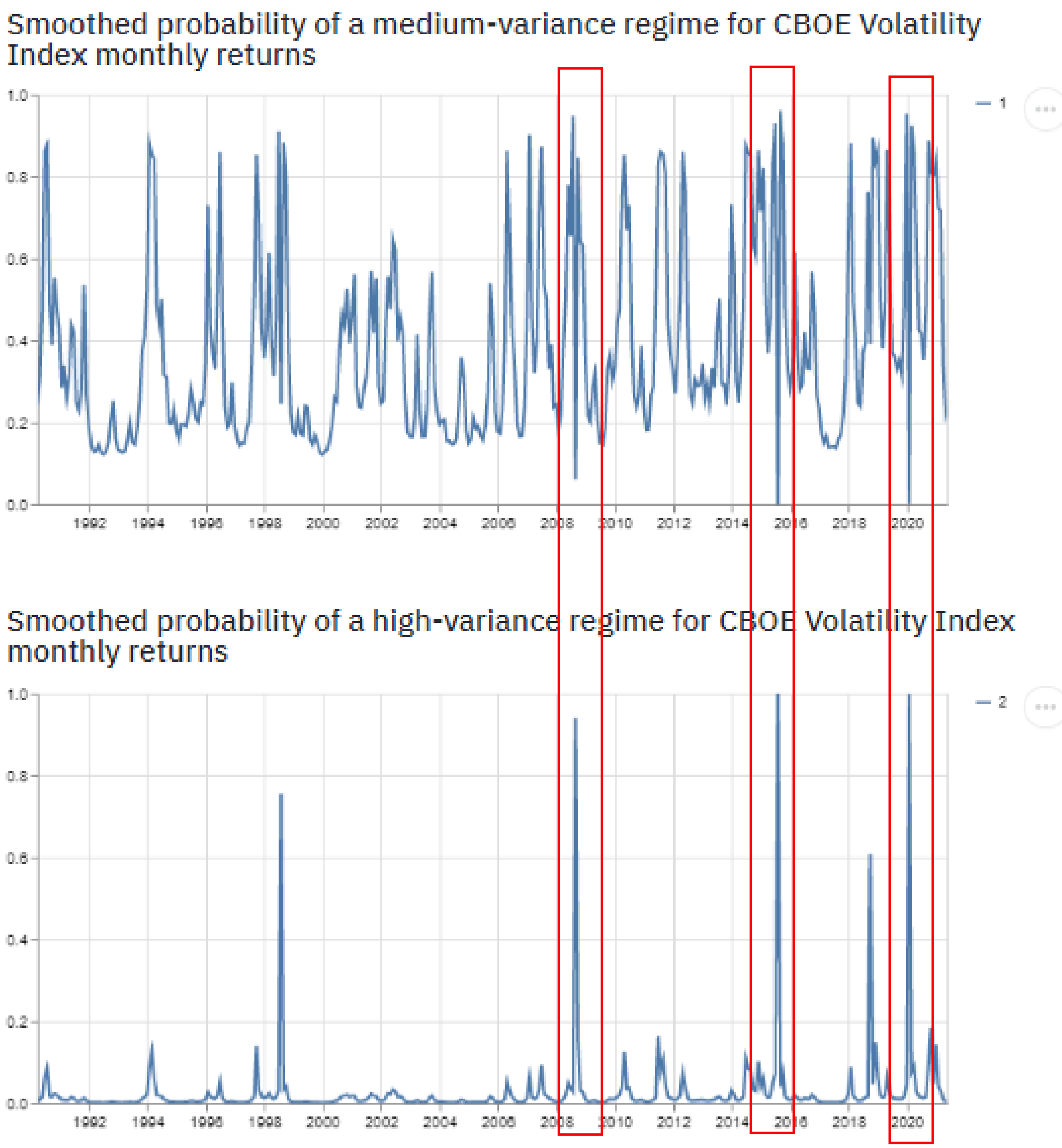


The first observation was looking at high-variance regime spikes.



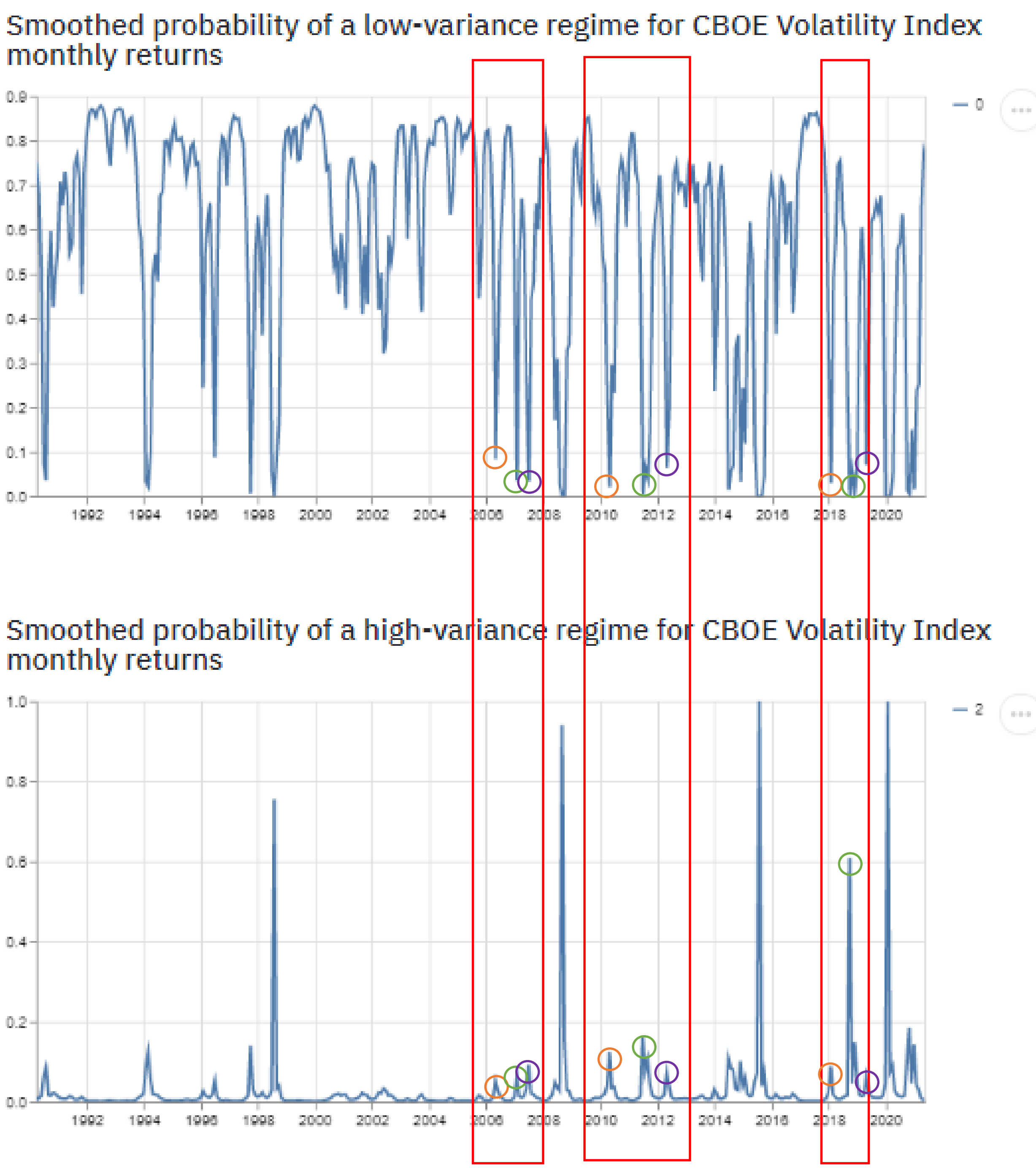
I noticed that there are some minor spikes prior to big spikes in high-variance regimes. This could possibly give lead to the idea that large spikes in high variance regimes don't occur by randomly they may be grouped with minor spikes.

The other thing that I had noticed was looking at the relationship between the two regimes. When looking at low-variance and high-variance we can see the change with large spikes.



We can see that the large spikes are associated with changes in the medium to high regime switches. This may give insight on the how these large spikes occur. It could be the case that the market switches from small to medium spikes which shows signs of more volatility and then they switch from medium to higher switches. It also shows that because most of the bigger spikes in high variance returns are associated with drops in the medium regimes there is already a *general consensus* of risk in the market.

When looking at low variance regime switches to high variance regimes switches we see.



This may mean that the large spikes only occur during transitions from medium to high variance, but more research is necessary. Something to investigate is the relationship with high-variance probabilities coming from either low-variance or medium-variance probabilities. Something to investigate is that a higher volatility may materialize to a bubble / recession if the transition comes from medium to high and it could be the case that small to high transitions will not materialize.

Ideas:

The initial goal of this was to apply this to left-tail hedging strategies. One possibility was looking at what happens before crashes. This still needs to be reviewed more by looking at the how the daily and weekly probabilities look for those specific dates. One idea would be to use the frequencies as lensing tools. It could be the case if we observe something on the monthly time series and then *zoom* in by looking at finer frequencies. Or it could be done vice versa going from daily to weekly to monthly.

Another idea is detecting whether higher variance regimes are market selloffs that will have rebounds, or they will materialize into bubbles. One approach would be detecting the medium to high variance changes, which looks like a sign of a bubble, and observing that scenario and then searching for the overvalued security or security class. The other method could be looking at value indicators (Buffet Indicator or Shiller PE CAPE indicator) and then linking medium to high variance changes to overvalue signals from those indicators. I could also make my own value indicators as well.

According to Black Swan Theory, Black Swans events contain a large amount of information within them. Looking at these smaller spikes (which technically are not Black Swans but sort of micro-Black Swans) they may contain important pieces of information within them. This could mean that looking into the spikes that are not associated to crashes to extract information from them.

This model could be used to predict left tails. Although this thinking opposes the theory behind modeling within Black Swan theory it seems like this model could provide an accurate predictor of Black Swans events. Even if the indicator had a *reliable* level of confidence in predicting Black Swans it probably could still be used, given that Universe (Taleb & Spitznagel) says that 95% of their left-tail hedges lose money (link here).

On the converse if you assume that financial models that dictate the market as well as the market itself does not have or promote antifragility then times of low variance would be times for finding cheaper left-tail hedges. There is some implications to this idea that lead to some problems. Problems that occur when using this model to find cheaper hedges include security type (inverse ETF, derivatives), duration on derivatives, and strike prices on those derivatives.

Areas of Concerns:

There is a large room for error. I break those areas into these categories.

- Programming errors – this could be errors within the statsmodels or within my code.
- Data errors – errors within the data that is getting pulled.
- Knowledge gaps – I do not fully understand markov chains, processes, or regime switching models to be fully confident in implementing a model. I also fully don't understand how to apply them in the python programming language let alone the statsmodels API.
- Statistical errors – this would be errors when drawing and making statistical inferences from the time series.

Another area of concern is the performing future analysis on this model. Markov regime switching models are a form of time series analysis. Now we are trying to analyze the time series of model that analyzes a time series. Possible next steps could be implementing more statsmodels tools or using another regime switching model, or continuous wavelet transform. But applying a second form of analysis may be overkill.

Another problem is that an indicator made from this may not be able to detect those high variance moments fast enough. But there are still applications of using machine learning to determine when a high variance probabilities will materialize to big selloffs.

Other uses:

- using it to for risk parity – volatility targeting models
- using it to stress test portfolios – determining what types of stress testing to use for a portfolio for a given timeline
- determining when rebalancing is necessary – if there are spikes that won't materialize then it probably isn't worth rebalancing and instead weathering
- building indicators – creating an indicator to detect changes in volatility or other financial time series
- applying this to different time series – market indices, interest rates, macroeconomic indicators, or portfolio

Diego Alvarez
Diego.alvarez@colorado.edu
LinkedIn | GitHub | Medium

Introduction:

I have recently built a web facing app (link here) that uses a repository I made in GitHub (link here, example here). The repo exclusively contains python code. The overall goal of this project is to create indicator like functions for volatility.

Methodology:

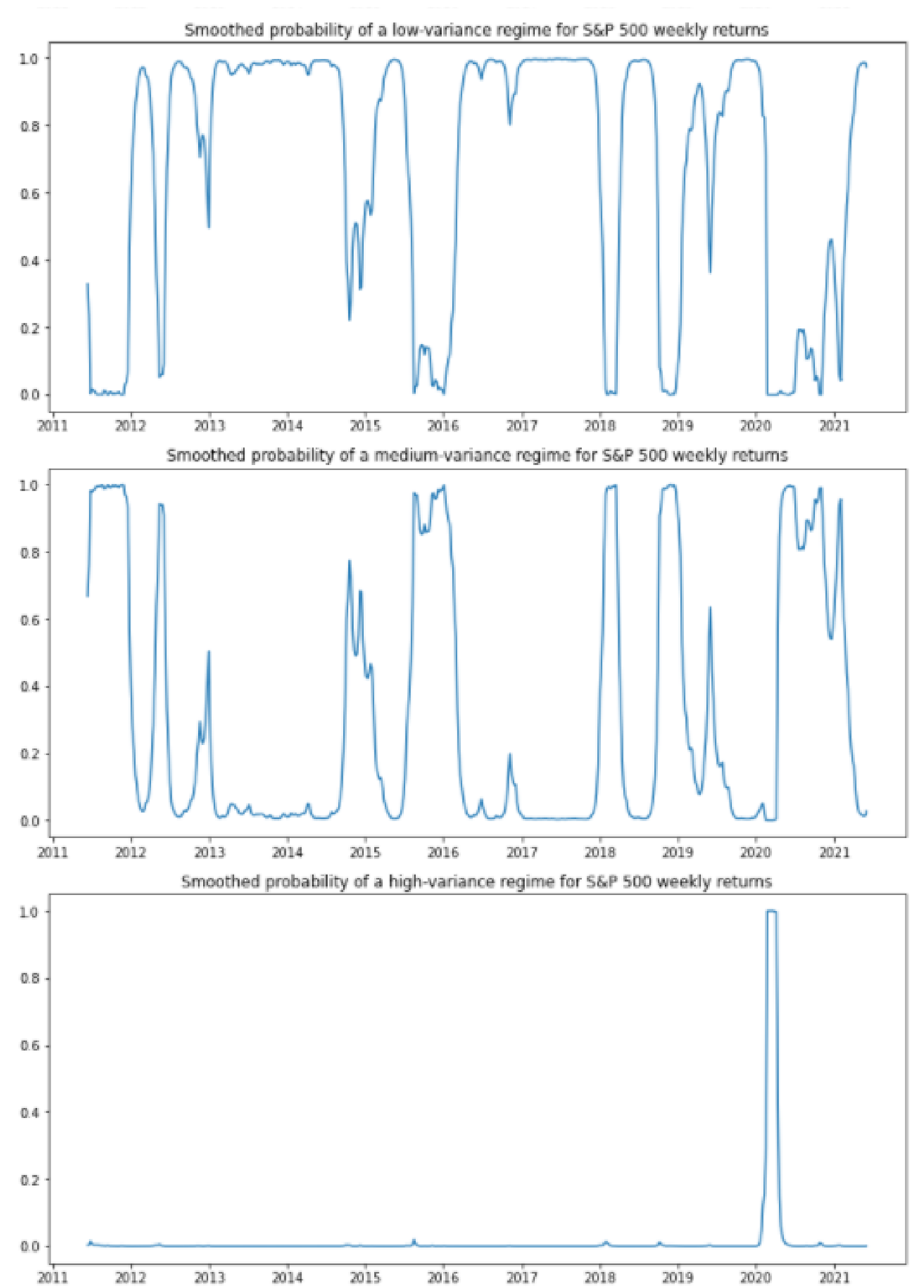
The data used in these experiments involves pulling data from yahoo finance by using the yfinance python API (see documentation here). The data was pulled at various time frames included (daily, weekly, monthly), and all of the prices are the Adjusted Close Price. The functions used for calculating the markov regime switching model is the python statsmodels API (see documentation here). The method used for finding the smoothed probability of the variance of returns for each regime is also built upon sample code from the statsmodels API (see 2nd example of the Kim, Nelson, and Startz (1998) Three-State Variance Switching link here). All of the code is accessible on the GitHub and all of the results are replicable using the Streamlit app.

Background:

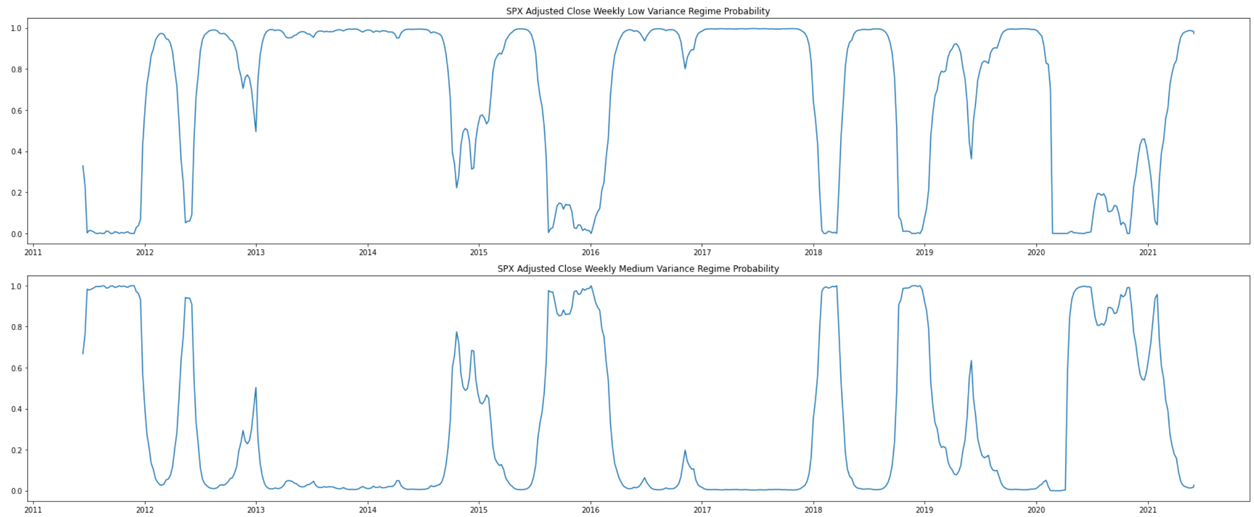
Please read the first writeup that I published about this topic (see link). To sum it up I'm looking at the creating Markov regime switching models that look at the variance of returns.

Findings:

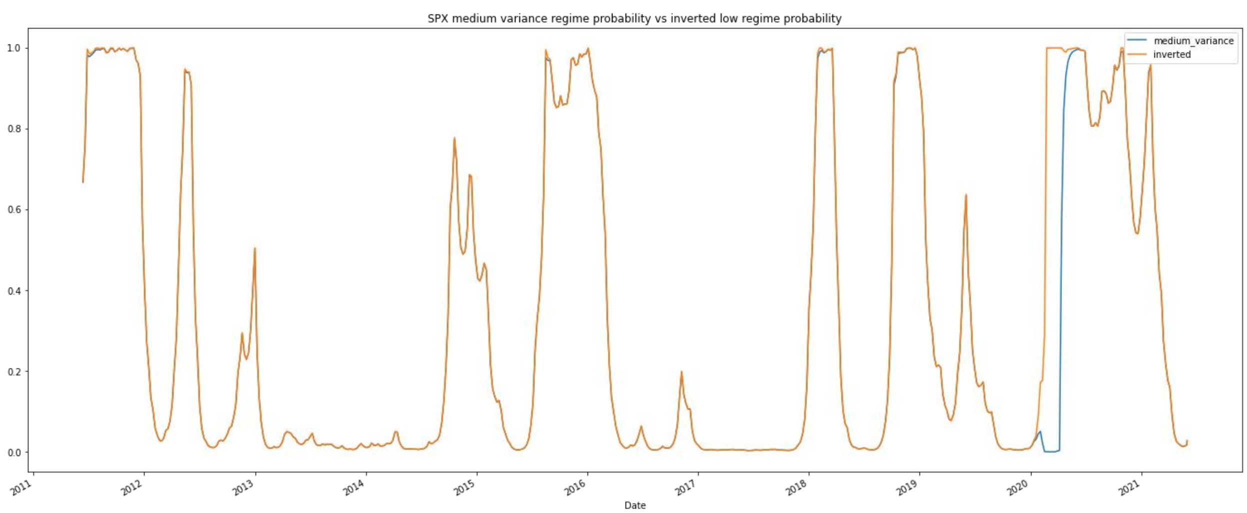
For this example, I pulled the S&P 500 weekly Adjusted Close price. I plotted the probability of Markov regimes. The regimes are determined by the variance of the returns.



Something that I noticed was that the medium and low regimes are almost the same time series but reflected across the x-axis.

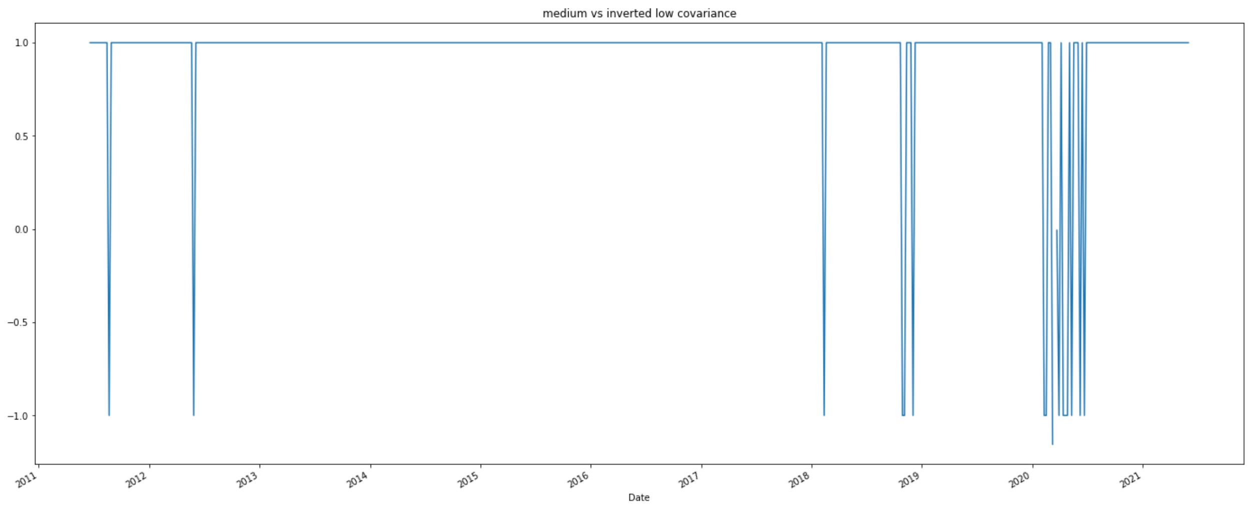


The first thing that I did was then invert one time series over the other.

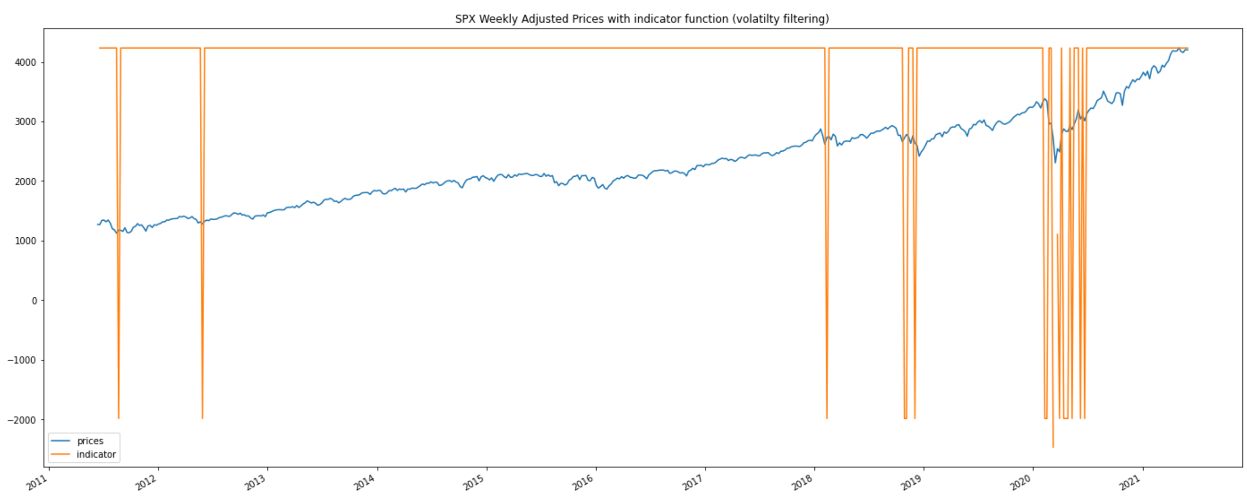


Then to make it into an indicator function I measured the correlation between the two-time series.

They will have almost perfect correlation until they break which will result in a spike in correlation. The correlation plot looks like.



Then if normalize the plot to the prices of the security and overlay this indicator function over the prices we get.



The streamlit app can output this graph for any security (left hand sidebar -> experiment, beta version is out there are still some bugs that need to be worked out)

Predictive Power:

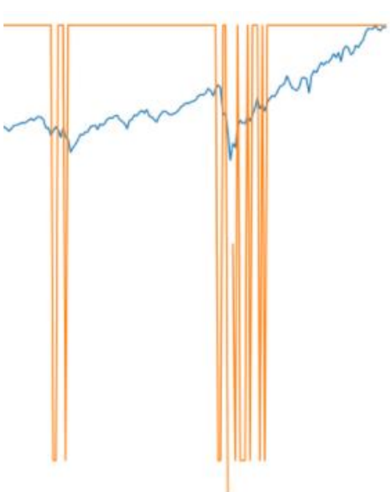
As seen, it looks like there is some applicability for trying to predict volatility. Although there needs to be more research into other time frames, different quoting frequencies, and with different types of securities. It may have some predictive power, but it may not be able to "predict" fast enough.

Using it as a risk measure:

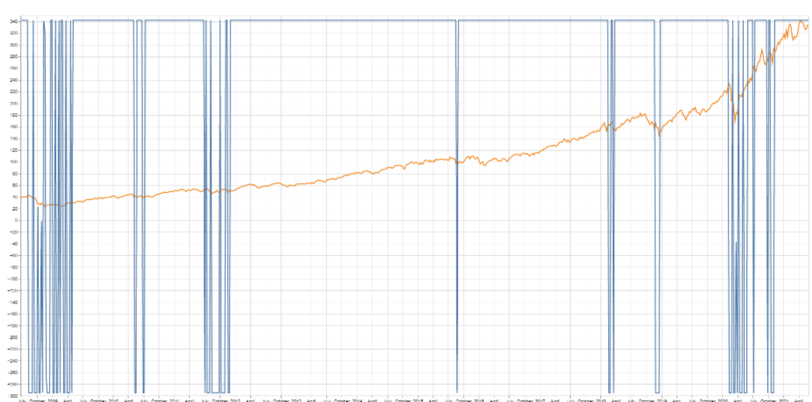
The application may be to run this and then "count" how many times the indicator function has shown signs. This could lead to a new risk measure of for these securities.

Looking into frequency of the indicators:

Another research vector would be to look into the how the frequency affects the price. We can see that with indicators that are closely packed together the price is volatile.



SPX example



QQQ Example

Inverting to different regimes.

Essentially, we are inverting the medium regime to low regime to get an indicator function for the high regime. In theory we should be able invert any to two regimes and measure correlation to get indicator functions for the third regime.

Areas of Concerns:

There is a large room for error. I break those areas into these categories.

- Programming errors – this could be errors within the statsmodels or within my code.
- Data errors – errors within the data that is getting pulled.
- Knowledge gaps – I do not fully understand markov chains, processes, or regime switching models to be fully confident in implementing a model. I also fully don't understand how to apply them in the python programming language let alone the statsmodels API.
- Statistical errors – this would be errors when drawing and making statistical inferences from the time series.

Another area of concern is the performing future analysis on this model. Markov regime switching models are a form of time series analysis. Now we are trying to analyze the time series of model that analyzes a time series. Possible next steps could be implementing more statsmodels tools or using another regime switching model, or continuous wavelet transform. But applying a second form of analysis may be overkill.

Diego Alvarez
Diego.alvarez@colorado.edu
LinkedIn | GitHub | Medium