# Crystal Ball Gazing: Mathematical models for chasing returns

Thomas P. Steele

June 17, 2025

## 0.1 Preface

At the end of 2019 I had started my first real job and for the first time in my life started making real money. As a 19 year old living at home with no meaningful expenses I quickly learned about the common vices of our world, alcohol and gambling. My beverage of choice was a single malt scotch whiskey, in particular Loch Lomond and Laphroaig, and my favourite method of gambling was on the stock market. On the night of the 2020 US election I wasn't celebrating or commiserating my political party, I was trading 500x leveraged CFDs on NSDQ100 and SPX500 and took my account from about 200USD to about 1800USD. I wasn't a complete fool, I withdrew enough money that I had completely covered what I had put in, so I was playing with the houses money, but, I had gained something else that night. Dangerous amounts of misplaced confidence. In the next few weeks I lost all of my winnings, then, started funding my account to keep playing until one night I realised I'd gambled away my entire last weeks pay. The anxiety of watching the number go down was crippling, and really made the lowest point in my life a lot lower. I cleaned myself up and got my shit together but that part of my story is not what got me here, no. I vowed revenge.

Every failure is a learning opportunity, and there were a lot of lessons learned. One, know your risk tolerance. You will lose when you lose your nerve and make decisions based on panic. Two, I don't have the nerve for day trading.

I followed lesson one, instead of gambling big on speculative mining companies and making 2-3x in a few weeks, instead, I put my money into ETFs and made respectable returns. Two I decided to give up my dream of being a day trader... and turned it into a dream of having a program to do it for me.

In 2021 I started university. In my second semester, MZB126, we learned about ODEs and it changed my world. I felt like I had just learned the why of all of the mathematics I had ever sat through and had the tools to understand everything, I felt like I was staring into the face of god themself. I wrote a physics engine in MATLAB since I had learned that in the same class, and made a lil rocket game where you flew around a rocket. I decided in this time too that computational physics is my future and just coded. I went through maybe 5 versions of that rocket game, made a 2D projection of an n-cube you could move around and play with and I learned how much you can learn from doing these sorts of self-motivated projects. I realised that if I practiced something, I would learn, it and become good at it.

Shortly after Christmas 2021 the rocket and the cubes were reaching a fairly polished state and it was time for a new project, I already knew what it would be. I had one learning objective and one goal; learn Python my second programming language, and, make a trading algorithm to take my revenge. I would start CBG (Crystal Ball Gazing).

I knew that I was far too emotional to day-trade. The simple fact is that If I felt something I would find technical indicators to support the decision I had already made, and then lose money. My goal was to create a program which would look at the market and decide an entry, stop-loss, and take-profit price and execute the trades on etoro. No emotions, only silicon and mathematics. I learned a lot in that first attempt, but frankly, the amount that I did not know meant that that first attempt was nothing more than a learning experience. My next major attempt came at the end of 2022 a year later when I did CAB420 Machine Learning, I implemented all sorts of models, but, I never-ever created a program I believed in.

There have been plenty of minor attempts after this whenever I had an idea or learned something new, but doing my final year of physics, then my honours of mathematics I simply did not have the time or mental capacity to tackle this project.

Until now.

After finishing my Honours I decided to look for work, and as a part of looking for work I decided that having some of my projects on GitHub would hopefully supplement my resumé neatly. Since I am looking for analytical roles including finance I decide that I should dig this project up for the world to see. I think I still have every major version of CBG, but my old code disgusts me, so, I essentially restarted from scratch.

And that's where we are today, that is what this is. Part complement to my resumé, part revenge for that money I lost.

The idea of this document is serve as a tour guide for the controlled chaos happening elsewhere. I am hoping to write down the ideas and mathematical derivations that underpin what I am doing, so hopefully another person could actually understand wtf is going on here. And maybe, just maybe, convince someone that I would be a valuable member of their team.

# Contents

# Chapter 1

# Mathematical Tools

## 1.1 Stochastic Models and Inference

A model in general takes some input and gives some output,

$$f(x) = y$$

A stochastic model is the same, however, every single time we evaluate the model we get a different answer.

To get a feel for stochastic modelling, we will start with a basic stochastic model. Suppose we are studying some phenomena and collect a bunch of observations. We will call the set of observations $Y$ and each observation $y_i$. Our goal is to make a stochastic model which mimics our set of observations,

$$y_i = \mu + \varepsilon_i$$

you can think of $\varepsilon$ as the *error*, so our model is that each of our observations has a value $\mu$ plus some error. For this type of model we must make an assumption on how our error is distributed, the easiest and often best assumption is that our error is normally distributed with mean 0 and some standard deviation $\sigma$, $\varepsilon\_i \sim \mathcal{N}(0, \sigma)$. Our model then becomes,

$$y_i \sim \mathcal{N}(\mu, \sigma),$$

meaning, our observations are randomly distributed by the normal distribution with mean $\mu$ and standard deviation $\sigma$. To actually lean anything from this model, we must first learn about the likelihood function.

The likelihood function is the total probability of our observations given our observations. What it tells us is the total probability of getting a set of observations given a model. The full form of the likelihood function is,

$$L\left(y_i \mid \theta\right) = \prod_{i=1}^{n} \pi\left(y_i \mid \{y_j\}_{j \neq i}^{n}, \theta\right). \tag{1.1}$$

Here $\theta$ is all of our model parameters, in this case $\mu$ and $\sigma$. A factor we must consider about our data is whether it is independent and identically distributed (iid) or not. iid data means that all of our observations are completely independent from one another, we will worry about non-iid data later. Here assuming our data is iid the likelihood becomes,

$$
\begin{aligned}
L\left(y_i \mid \theta\right) &= \prod_{i=1}^{n} \pi\left(y_i \mid \theta\right), \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(y_i-\mu\right)^2\right), \\
&= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu\right)^2\right).
\end{aligned}
$$

The parameters which are the most likely to give us our observations maximises the likelihood function, so to infer the value of our parmeters, we can find for what values is the likelihood function maximised. Before we get started on that, let's talk about the log-likelihood. The log-likelihood is exactly the same, but, wildly easier to differentiate,

$$
\begin{aligned}
\ell\left(Y \mid \theta\right) &= \ln(L\left(Y \mid \theta\right)) \\
&= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2
\end{aligned}
$$

To find the best value for $\mu$ we solve,

$$
\begin{aligned}
0 &= \frac{\partial \ell\left(Y \mid \theta\right)}{\partial \mu}, \\
0 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu), \\
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} y_i,
\end{aligned}
$$

and for $\sigma$,

$$
\begin{aligned}
0 &= \frac{\partial \ell\left(Y \mid \theta\right)}{\partial \sigma}, \\
0 &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i-\mu)^2, \\
\hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2}.
\end{aligned}
$$

We give our estimates $\hat{\mu}$ and $\hat{\sigma}$ to signify that these are estimated values. If these formulas look familiar, they should, these are the mean and standard deviation formulas, the mean in particular you probably use far more often than you realise.

### *1.1.1 Example: Linear Regression*

Linear regressions have a lot of different forms, that are in the end all special cases of the same fundamental idea, that is, there is a linear relationship between some predicting factors $\mathbf{x}$ and some resulting outcome $\mathbf{y}$. Suppose there is $p$ factors which influence the outcome of our observations. For each experiment $i \in [1, \ldots, n]$ we an outcome $y_i$, and factors which we think predicts $y_i$ $\mathbf{x}_i$, where,

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$$

Our stochastic model is a weighted linear combination of the predictors $\mathbf{x}_i$ to predict the outcome $y_i$,

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + \varepsilon_i,$$

Since we have a linear system, we can use the tools of linear algebra to clean it up,

$$\begin{aligned} y_i &= \mathbf{x}_i^{\intercal}\beta + \varepsilon_i \\ \mathbf{x}_i &= \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma). \end{aligned}$$

Here we have redefined $\mathbf{x}_i$ to have 1 as its first element, this is so that the constant $\beta_{i0}$ is not interfered with. Next we build the likelihood function and assume that, as we essentially do for all linear regressions, that our data is iid and normally distributed,

$$\begin{aligned} y_i &\sim \mathcal{N}(\mathbf{x}_i^{\intercal}\beta, \sigma), \\ L(\mathbf{Y} \mid \theta) &= \prod_{i=1}^{n} (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \mathbf{x}_i^{\intercal}\beta\right)^2\right) \\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\mathbf{y}_i - \mathbf{x}_i^{\intercal}\beta\right)^2\right). \end{aligned}$$

Our observations $y$ are distributed by the multivariate normal distribution with mean $\mathbf{x}_i^\intercal \beta$ and with standard deviation $\sigma$. We can further simplify the likelihood by defining,

$$
\begin{aligned}
\mathbf{X} &= \begin{bmatrix} \mathbf{x}_1, & \mathbf{x}_2, & \ldots, & \mathbf{x}_n \end{bmatrix}^\intercal, \\
\mathbf{Y} &= \begin{bmatrix} y_1, & y_2, & \ldots, & y_n \end{bmatrix}^\intercal, \\
\mathbf{Y} &= \mathbf{X}\beta + \varepsilon, \\
L\left(\mathbf{Y} \mid \theta\right) &= (2\pi)^{-\frac{d}{2}} \sigma^{-n} \exp\left( -\frac{1}{2\sigma^2} \left(\mathbf{Y} - \mathbf{X}\beta\right)^\intercal \left(\mathbf{Y} - \mathbf{X}\beta\right) \right).
\end{aligned}
$$

Now, using some careful multivariate calculus we can find $\hat{\beta}$ and $\hat{\sigma}$ using maximum likelihood estimation,

$$
\begin{aligned}
\frac{\partial \ell\left(\mathbf{Y} \mid \theta\right)}{\partial \beta} &= \sigma^{-2} \mathbf{X}^\intercal \left(\mathbf{Y} - \mathbf{X}\beta\right) = 0 \\
\hat{\beta} &= \left(\mathbf{X}^\intercal \mathbf{X}\right)^{-1} \mathbf{X}^\intercal \mathbf{Y} \\
\frac{\partial \ell\left(\mathbf{Y} \mid \theta\right)}{\partial \sigma} &= -\frac{n}{\sigma} + \sigma^{-3} \left(\mathbf{Y} - \mathbf{X}\beta\right)^\intercal \left(\mathbf{Y} - \mathbf{X}\beta\right) = 0 \\
\hat{\sigma} &= \sqrt{\frac{1}{n} \left(\mathbf{Y} - \mathbf{X}\beta\right)^\intercal \left(\mathbf{Y} - \mathbf{X}\beta\right)}
\end{aligned}
$$

## 1.2 Bayes Theorem and Bayesian Inference

Summation rule

$$\sum_{i=1}^{n} \Pr\left(x, y_i\right) = \Pr\left(x\right) \tag{1.2}$$

Product Rule

$$\Pr\left(A \cap B\right) = \Pr\left(A\right) \Pr\left(B \mid A\right) \tag{1.3}$$

Bayes theorem comes from thinking about the product rule being commutative,

$$\Pr\left(A \cap B\right) = \Pr\left(B \cap A\right) \tag{1.4}$$

$$\Pr\left(A\right) \Pr\left(B \mid A\right) = \Pr\left(B\right) \Pr\left(A \mid B\right) \tag{1.5}$$

$$\Pr\left(A \mid B\right) = \frac{\Pr\left(A\right) \Pr\left(B \mid A\right)}{\Pr\left(B\right)} \tag{1.6}$$

The usual form of Bayes' theorem we will see when doing Bayesian Inference is,

$$\pi\left(\theta \mid X\right) = \frac{L\left(X \mid \theta\right) \pi\left(\theta\right)}{\int_{\theta \in \Theta} L\left(X \mid \theta\right) \pi\left(\theta\right) \, \mathrm{d}\theta} \tag{1.7}$$

The denominator comes from Bayes Theorem,

$$\begin{aligned}\pi(X) &= \int_{\theta\in\Theta} \pi(X\cap\theta)\,\mathrm{d}\theta \\ \pi(X\cap\theta) &= L(X\mid\theta)\,\pi(\theta) \\ \pi(X) &= \int_{\theta\in\Theta} L(X\mid\theta)\,\pi(\theta)\,\mathrm{d}\theta\end{aligned}$$

### *1.2.1 Bayesian Linear Regression*

We use the same example here as in 1.1.1, we have a set of $n$ vectors of predictors $\mathbf{x}$ predicting an outcome $y$ which we assume are normally distributed,

$$\begin{aligned}L(\mathbf{Y}\mid\theta) &= \prod_{i=1}^{n}(2\pi)^{-\frac{1}{2}}\sigma^{-1}\exp\left(-\frac{1}{2\sigma^2}\left(y_i-\mathbf{x}_i^{\intercal}\beta\right)^2\right) \\ &= (2\pi)^{-\frac{d}{2}}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)\right)\end{aligned}$$

For Bayesian linear regressions we typically use the precision $\tau$ instead of the standard deviation $\sigma$ where $\tau=\sigma^{-2}$. This typically makes our lives just a little bit easier. The likelihood then is,

$$(2\pi)^{-\frac{d}{2}}\tau^{-\frac{n}{2}}\exp\left(-\frac{\tau}{2}\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)\right).$$

After determining the likelihood, the next step of bayesian inference is to define the prior. The prior distribution is a pre-determined probability distribution of the parameter values, here $\beta$ and $\tau$. We will use a *conjugate prior* basically, making this a lot easier for ourselves. The prior distribution on our parameter values should reflect what we know *a priori*, however, not to be too restrictive that it unduly influences the outcome. The weights $\beta\in\mathbb{R}$ so the most logical choice is the multivariate normal distribution $\mathrm{Normal}\left(\beta\,\middle|\,\beta_0,\lambda^{-1}\right)$ ($\beta_0$ is the prior mean of $\beta$ and $\lambda^{-1}$ is the prior covariance) and the precision $\tau\in\mathbb{R}_+$ so it must be distributed by something which is in $\mathbb{R}_+$ and a gamma distribution is a logical choice $\mathrm{Gamma}(\tau\mid a,b)$.

Our can be a simple mixture of these distributions $\pi(\beta,\tau\mid\beta_0,\lambda,a,b)=\mathrm{Normal}\left(\beta\,\middle|\,\beta_0,\lambda^{-1}\right)\mathrm{Gamma}$ however, it is actually better to use the *Normal-gamma* distribution. The normal-gamma is largely the same, however, $\beta$ is distributed by $\mathrm{Normal}\left(\beta\,\middle|\,\beta_0,(\lambda\tau)^{-1}\right)$, i.e. we choose $\beta$ conditional on the precision. In truth this is simply done for mathematical convenience since the normal gamma distribution is the conjugate prior to the normal distribution where the mean and precision are unknown.

$$\begin{aligned}\mathrm{NormalGamma}(\beta,\tau\mid\beta_0,\lambda,a,b) &= (2\pi)^{-\frac{d}{2}}\left|\lambda\tau\right|^{\frac{1}{2}}\exp\left(-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right)\frac{b^a}{\Gamma(a)}\tau^{a-1}e^{-b\tau} \\ &= \frac{b^a\sqrt{|\lambda|}}{(2\pi)^{-\frac{d}{2}}}\tau^{a-\frac{2-d}{2}}e^{-b\tau}\exp\left(-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right)\end{aligned}$$

Frankly, this just *is* a complicated looking distribution. Now, we can determine our posterior,

$$
\begin{aligned}
\pi\left(\theta \mid \mathbf{Y}\right) &= \frac{L\left(\mathbf{Y} \mid \theta\right)\pi\left(\theta\right)}{\int_{\theta\in\Theta} L\left(\mathbf{Y} \mid \theta\right)\pi\left(\theta\right)\,\mathrm{d}\theta} \\
&= \frac{(2\pi)^{-\frac{d}{2}}\tau^{\frac{n}{2}}\exp\left(-\frac{\tau}{2}\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)\right)\frac{b^a\sqrt{|\lambda|}}{(2\pi)^{-\frac{d}{2}}}\tau^{a-\frac{2-d}{2}}e^{-b\tau}\exp\left(-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right)}{\int_0^{\infty}\int_{-\infty}^{\infty}\frac{b^a\sqrt{|\lambda|}}{(2\pi)^{-\frac{d}{2}}}\tau^{a-\frac{2-d}{2}}e^{-b\tau}\exp\left(-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right)\,\mathrm{d}\beta\,\mathrm{d}\tau}
\end{aligned}
$$

This is absolutely heinous looking to work with, so we will start again using the Bayesians favourite symbol, proportional to,

$$
\begin{aligned}
\pi\left(\theta \mid \mathbf{Y}\right) &\propto L\left(\mathbf{Y} \mid \theta\right)\pi\left(\theta\right) \\
&\propto (2\pi)^{-\frac{d}{2}}\tau^{\frac{n}{2}}\exp\left(-\frac{\tau}{2}\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)\right)\frac{b^a\sqrt{|\lambda|}}{(2\pi)^{-\frac{d}{2}}}\tau^{a-\frac{2-d}{2}}e^{-b\tau}\exp\left(-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right)
\end{aligned}
$$

Our goal is to get this into the form of a known probability density. Since we are using conjugate priors it will be the normal-gamma distribution. First we drop every multiplicative factor that is not our data or parameters,

$$
\pi\left(\theta \mid \mathbf{Y}\right) \propto \tau^{a+\frac{n}{2}-\frac{2-d}{2}}e^{-b\tau}\exp\left(-\frac{\tau}{2}\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)-\frac{\tau}{2}\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right)\right),
$$

next we complete the square such that $\beta$ is the subject,

$$
\pi\left(\theta \mid \mathbf{Y}\right) \propto \tau^{a+\frac{n}{2}-\frac{2-d}{2}}e^{-b\tau}\exp(\left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)+\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right))^{-\frac{\tau}{2}},
$$

$$
\begin{aligned}
&= \left(\mathbf{Y}-\mathbf{X}\beta\right)^{\intercal}\left(\mathbf{Y}-\mathbf{X}\beta\right)+\left(\beta-\beta_0\right)^{\intercal}\lambda\left(\beta-\beta_0\right) \\
&= \mathbf{Y}^{\intercal}\mathbf{Y}-Y^{\intercal}\mathbf{X}\beta-\beta^{\intercal}\mathbf{X}^{\intercal}\mathbf{Y}+\beta^{T}\mathbf{X}^{\intercal}\mathbf{X}\beta+\beta^{\intercal}\lambda\beta-\beta^{\intercal}\lambda\beta_0-\beta_0^{\intercal}\lambda\beta+\beta_0^{\intercal}\lambda\beta_0 \\
&= \beta^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)\beta-\left(Y^{\intercal}\mathbf{X}+\beta_0^{\intercal}\lambda\right)\beta-\beta^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)+\mathbf{Y}^{\intercal}\mathbf{Y}+\beta_0^{\intercal}\lambda\beta_0 \\
&= \left(\beta-\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)\right)^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)\left(\beta-\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)\right)+c \\
c &= \mathbf{Y}^{\intercal}\mathbf{Y}+\beta_0^{\intercal}\lambda\beta_0-\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)
\end{aligned}
$$

$$
\begin{aligned}
\pi\left(\theta \mid \mathbf{Y}\right) \propto{} & \tau^{a+\frac{n}{2}-\frac{2-d}{2}}e^{-\tau\left(b+\frac{1}{2}\left(\mathbf{Y}^{\intercal}\mathbf{Y}+\beta_0^{\intercal}\lambda\beta_0-\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)\right)\right)}\dots \\
& \exp\left(\frac{\tau}{2}\left(\beta-\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)\right)^{\intercal}\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)\left(\beta-\left(\mathbf{X}^{\intercal}\mathbf{X}+\lambda\right)^{-1}\left(\mathbf{X}^{\intercal}\mathbf{Y}+\lambda\beta_0^{\intercal}\right)\right)\right),
\end{aligned}
$$

Finally, our posterior is a normal-gamma distribution, where the marginal posterior distributions of $\tau$ and $\beta$ is

$$\tau \sim \text{Gamma}\left(\tau \,\middle|\, a + \frac{n}{2}, b + \frac{1}{2}\left(\mathbf{Y}^\intercal\mathbf{Y} - \left(\mathbf{X}^\intercal\mathbf{Y} + \lambda\beta_0^\intercal\right)^\intercal \left(\mathbf{X}^\intercal\mathbf{X} + \lambda\right)^{-1} \left(\mathbf{X}^\intercal\mathbf{Y} + \lambda\beta_0^\intercal\right)\right)\right)$$
$$\beta \sim \text{Normal}\left(\beta \,\middle|\, \left(\mathbf{X}^\intercal\mathbf{X} + \lambda\right)^{-1} \left(\mathbf{X}^\intercal\mathbf{Y} + \lambda\beta_0^\intercal\right), \tau^{-1}\left(\mathbf{X}^\intercal\mathbf{X} + \lambda\right)^{-1}\right).$$

I think quite clearly the Bayesian linear regression involves more work, however, it gives us in return a more powerful solution, such as credible intervals which can help us more meaningfully reject parameters.

# Chapter 2

# Financial Random Walks

## 2.1 Discrete Time Continous Space Financial Random Walks

If you look on the wikipdeia page financial random walks you will be greeted with the following equation,

$$S_{t+1} = S_t + \mu \Delta t S_t + \sigma \sqrt{\Delta t} S_t Y_i. \tag{2.1}$$

This is a fairly typical random walk, except that our drift depends one our position $S_t$. If we factor $S_t$ we find,

$$S_{t+1} = S_t \left( 1 + \mu \Delta t + \sigma \sqrt{\Delta t} Y_i \right).$$

This gives us a recurrence relationship, thus we can deffine,

$$S_{t+n} = S_t \prod_{i=1}^{n} \left( 1 + \mu \Delta t + \sigma W_i \right),$$

taking the log we find,

$$\ln(S_{t+n}) = \ln(S_t) + \sum_{i=1}^{n} \ln(1 + \mu \Delta t + \sigma W_i).$$

To get to where I want to go next, we have two paths, by small value approximation, or by argument. I choose to argue. My argument is, the original random walk from wikipedia Equation (2.1) is flawed. We cannot have negative prices, this is an inutuive fact, however, this equation is $\Delta t$ is large enough will give negative prices, the proper error here should not be able to make our prices negative. The log of the price however, negative values are fair game. I propose therefore this random walk,

$$\ln(S_{t+n}) = \ln(S_t) + \sum_{i=1}^{n} \mu \Delta t + \sigma W_t. \tag{2.2}$$

#### *2.1.1 Time Series Models*

*The AR Model*

## 2.2 Continuous Time Continous Space Financial Random Walks

Equation (2.2) is a discrete time continuous space random walk. By taking the limit $\Delta t \to \mathrm{d}t$ we get a continuous time continuous space random walk,

$$\ln(S(t + \mathrm{d}t)) = \ln(S_t) + \mu \, \mathrm{d}t + \sigma \, \mathrm{d}W$$