Model Validation Report

Ezra Stuart

October 2025

Chapter 1

Introduction

This is a Vasicek model, trained on 10 year US Treasury Yields in order to be used in the context of pension funding ratios.

The implementation of the model is very good, and despite being a bit too simplistic to capture some of the short term dynamics, it is stable enough to forcast rates into the future. Despite this, the proposed application to funding ratios is extremely inappropriate as the model is, 1 trained on US rates, not UK ones and 2 not nearly enough to capture the full complexity of a pension portfolio and its dynamics.

I suggest that for this model to be useful in application to pension funding ratios:

- The documentation must be heavily cleaned up
- Be retrained on UK gilts with a range of maturities
- Be added to pipeline and used in conjucntion with other models to accurately represent the full complexity of a pensioin portfolio

Chapter 2

Data

Data Quality

The overall data quality that this model is trained on is quite good

- The 10 Year US Treasury Yields, were pulled from the Federal Reserve Economic Data, a very reliable source.
- There are 15 years in the train set ($\sim$ 180 data points).
- **However,** only 5 years of test data ($\sim$ 70 datapoints), which is straight from the Covid crash to today, not very representative of rates in the future.

Representativeness

For the specific use in UK pensions, this data is not the most representative. Although rates in the US and UK are indeed inextricably linked, for the intended use case, I would suggest UK Gilt yields instead of US Treasury yields.

The data does in fact cover the full rate cycle, and crisis periods (2008 in training and 2020 in test)

2.1 Model Applicability

Stationarity Tests

The Augmented Dickey Fuller (ADF), Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and Zivot-Andrews tests can be used to investigate the training time series is stationary, as the Vasicek model, which assumes stationarity is used. Meaning that the series has, constant mean, constant volatility and displays no seasonality. The results of these tests are:

- ADF: -1.8, $p = 0.38 > 0.05$ - Accept null hypothesis, insignificant evidence to assume that the series is stationary
- Zivot-Andrews: -4.8, $p = 0.10 > 0.05$ - Accept null hypothesis, insignificant evidence to assume that the series is stationary
- KPSS: 1.36, $p < 0.01 < 0.05$ - Reject null hypothesis, significant evidence to suggest that the time series is not stationary

I believe that this is because the training data, spans two different monetary regimes (pre and post 2008 crash) and there appears to be some degree of seasonality in that section of the series.

Upon concatination (2010 - 2020), the results of these tests are:

- ADF: -2.70, $p = 0.074 > 0.05$ - Accept null hypothesis, insignificant evidence to assume that the series is stationary
- Zivot-Andrews: -3.3, $p = 0.88 > 0.05$ - Accept null hypothesis, insignificant evidence to assume that the series is stationary
- KPSS: 0.255, $p > 0.1 > 0.05$ - Accept null hypothesis, insignificant evidence to suggest that the time series is not stationary

Even after this reduction, only one of the three tests reaches a stationarity conclusion, therefore we have significant evidence that the current Vasicek model is unfit for the purpose of modelling 10 year treasury yields.

First Difference

Applying these tests to the first difference of the training series we find:

- ADF: -6.50, $p = 1.16e^{-8} < 0.05$ - Reject null hypothesis, significant evidence to suggest that the series is stationary
- Zivot-Andrews: -6.9, $p = 0.00098 < 0.05$ - Reject null hypothesis, significant evidence to suggest that the series is stationary
- KPSS: 0.04, $p > 0.1 > 0.05$ - Accept null hypothesis, insignificant evidence to suggest that the time series is not stationary

These results suggest that the training period is best modelled by a random walk with drift. However, I do not believe that this is the best approach to modelling interest rates in a long term scenario, as the negative drift will compound over longer time horizons.

Variance Ratio

In order to test this Random Walk hypothesis I will carry out a Variance ratio test on the first difference of the training data:

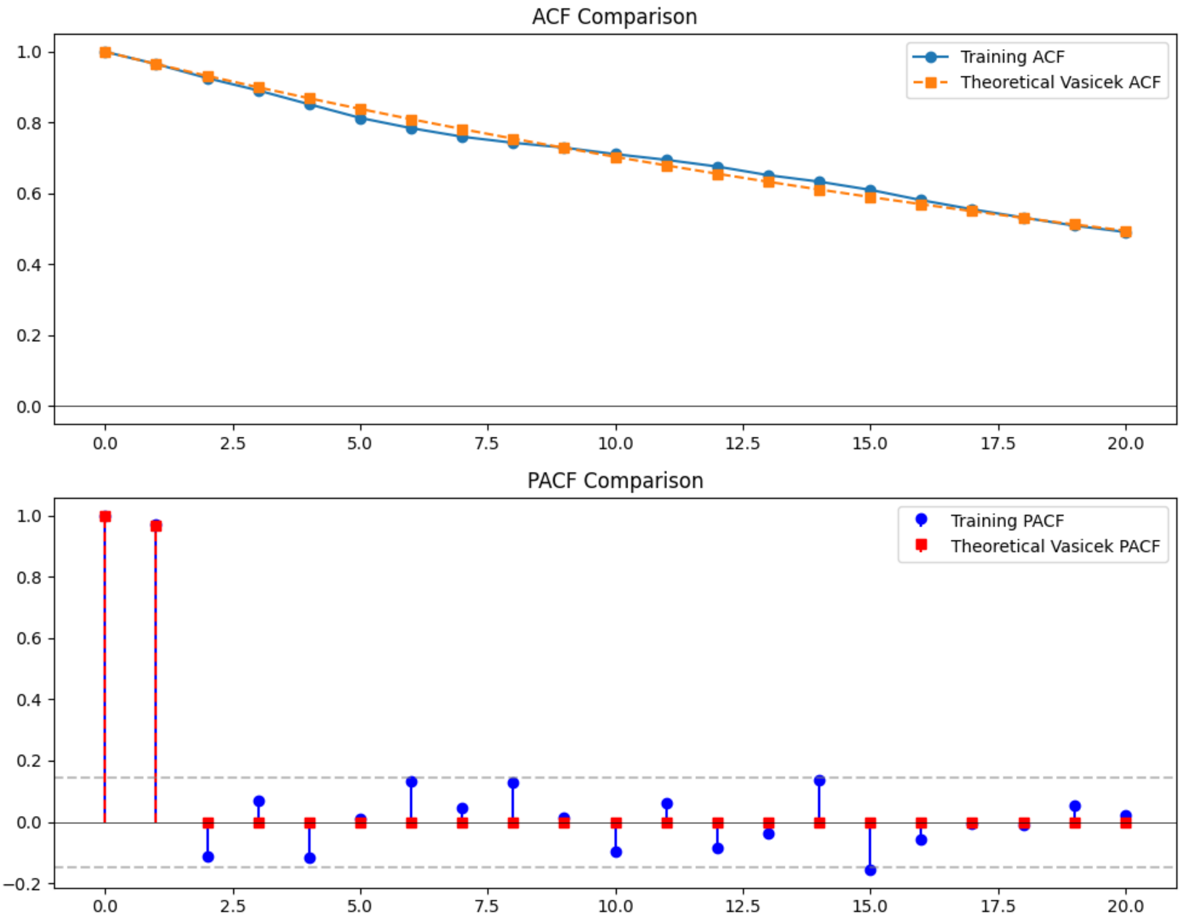
$\mathcal{H}_\prime$: Series follows a random walk
$\mathcal{H}_\prime$: Series does not follow a random walk.

Lag	Test Statistic	p-value
2	-4.4	0
5	-3.6	0
10	-3.0	0.003
20	-2.3	0.022

In conclusion, we reject $\mathcal{H}_0$ as there is definitely enough evidence to suggest that the series does not follow a random walk, and the negative test statistic suggests that either the series DOES show mean-reverting properties, or a Geometric Brownian Motion (GBM), whihc is 'slowing down'.

Auto-Correlation

For further analysis, I compared the AutoCorrelation (ACF) and Partial-Autocorrelation of this data against those of a theoretical Vasicek distribution, finding:



The ACF structure of our training data fits very nicely with that of a theoretical Vasicek, and the PACF structure only has a few lags where the differences are close to significant.

Chapter 3

Model Theory

3.1 Implementation

Solution

The choice of a Vasicek model in this case is quite strong, however the use of the Euler-Maruyama discretisation to solve the SDE is not reccomended. The vasicek model is an Ornstein-Uhlenbeck process, meaning that it can be solved analytically with little extra effort:

$$\begin{aligned} dr_t &= \kappa(\mu - r_t)dt + \sigma dW_t \\ dr_t + \kappa r_t dt &= \kappa\mu dt + \sigma dW_t \end{aligned} \tag{3.1}$$

Observing that;

$$d\left(e^{\kappa t} r_t\right) = e^{\kappa t} dr_t + \kappa e^{\kappa t} r_t dt$$

we can multiply through by $e^{\kappa t}$:

$$\begin{aligned} e^{\kappa t} dr_t + e^{\kappa t} \kappa r_t dt &= e^{\kappa t} \kappa\mu dt + e^{\kappa t} \sigma dW_t \\ \int_0^T d\left(e^{\kappa t} r_t\right) &= \int_0^T e^{\kappa t} \kappa\mu dt + \int_0^T e^{\kappa t} \sigma dW_t \end{aligned} \tag{3.2}$$

and evaluate the integrals:

$$\begin{aligned} e^{\kappa T} r_T - r_0 &= e^{\kappa T} \mu - \mu + \sigma \int_0^T e^{\kappa t} dW_t \\ r_T &= e^{-\kappa T} r_0 + \mu\left(1 - e^{-\kappa T}\right) + \sigma \int_0^T e^{-\kappa(T-t)} dW_t \end{aligned} \tag{3.3}$$

In order to find the analytic solution at time T, for this series, r_T

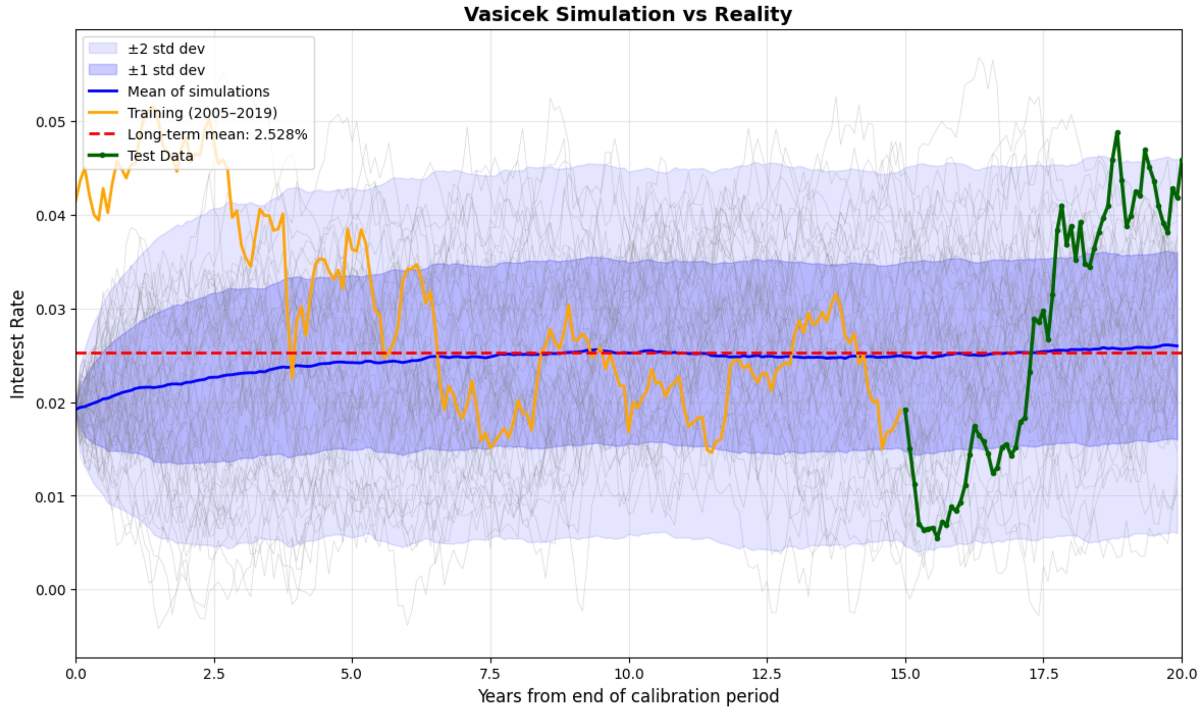
Parameters

The Maximum Likelihood Estimate method is used in order to calibrate the parameters, however, it assumes the Euler-Maruyama transition densities, instead of the correct Ornstein-Uhlenbeck ones.

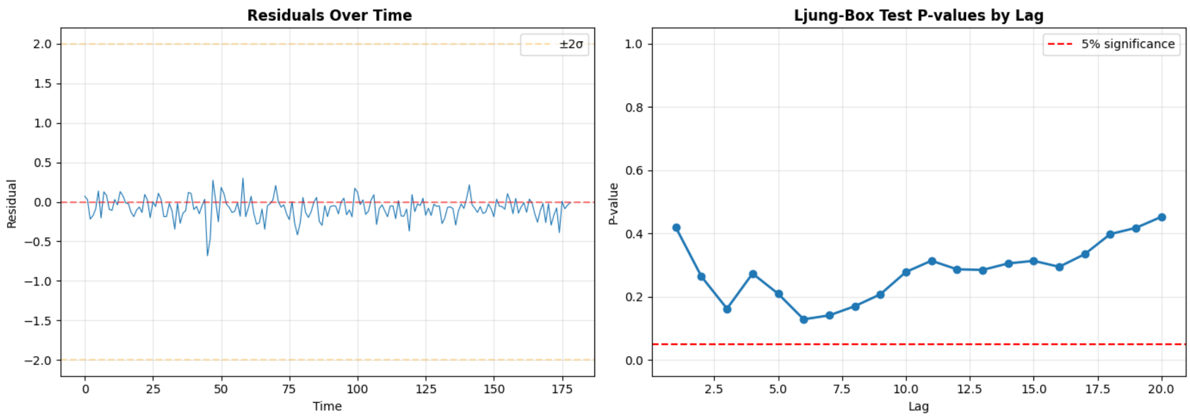
The calculated parameters are:

- $\mu = 2.5\%$
- $\kappa = 0.35$
- $\sigma = 0.8\%$

These parameters look reasonable as the historical training rates seem to oscillate about this long-term mean, as pictured below, and the reversion speed and volatility are inline with reasonable judgement.



Applying the Ljung-Box test, comparing the training rates with a theoretical Vasicek model, we find:



There is no significance to suggest that the residuals are not independently distributed, suggesting a Vasicek with these parameters is indeed a good fit

3.2 Challenger Models

The Vasicek model allows for negative rates, which has never been observed for 10 year US treasury yields. One model whihc prevents negative rate is the Cox-Ingersol-Ross (CIR) model:

$$dr_t = \kappa_C(\mu_C - r_t)dt + \sqrt{r_t}\ \sigma_C dW_t \tag{3.4}$$

Geometric Brownian Motion

As discussed in the data section, the first difference of training set suggests that a Geometric Brownian Motion Model could be used to better capture rate dynamics. A time series undergoing Geometric Browniann Motion obeys the SDE:

$$dr_t = vdt + \sigma_G dW_t \tag{3.5}$$

To solve this SDE, we can take the logarithm of r_t and apply Ito's Lemma;

$$\begin{aligned} dX_t = d(lnr_t) &= (v - \frac{1}{2}\sigma_G^2)dt + \sigma_G dW_t \\ \int_0^t dX_t &= (v - \frac{1}{2}\sigma_G^2)\int_0^t dt + \int_0^t \sigma_G dW_t \\ X_t - X_0 &= (v - \frac{1}{2}\sigma_G^2)t + \sigma_G W_t \end{aligned} \tag{3.6}$$

Therefore,

$$\begin{aligned} ln(r_t) &= ln(r_0) + (v - \frac{1}{2}\sigma_G^2)t + \sigma_G W_t \\ r_t &= r_0 \exp\left[(v - \frac{1}{2}\sigma_G^2)t + \sigma_G W_t\right] \end{aligned} \tag{3.7}$$

From the logartihmic relation, we can deduce the distribution of r_t.
Noticing that $W_t \sim \mathcal{N}(0,t)$ and the other terms are all deterministic, we can see that $\ln(r_t)$ is normally distributed and therefore r_t follows the lognormal, or more specifically:

$$\begin{aligned} \ln(r_t) &\sim \mathcal{N}\left(ln(r_0) + \left(v - \frac{1}{2}\sigma_G^2\right)t,\ \sigma_G^2 t\right) \\ r_t &\sim LogNormal\left(ln(r_0) + \left(v - \frac{1}{2}\sigma_G^2\right)t,\ \sigma_G^2 t\right) \end{aligned} \tag{3.8}$$

Knowing this, we can check to see if the training data follows the lognormal distribution by casrrying out a Kolmononogorv-Smirnov test comparign it to the log-normal:

Chapter 4

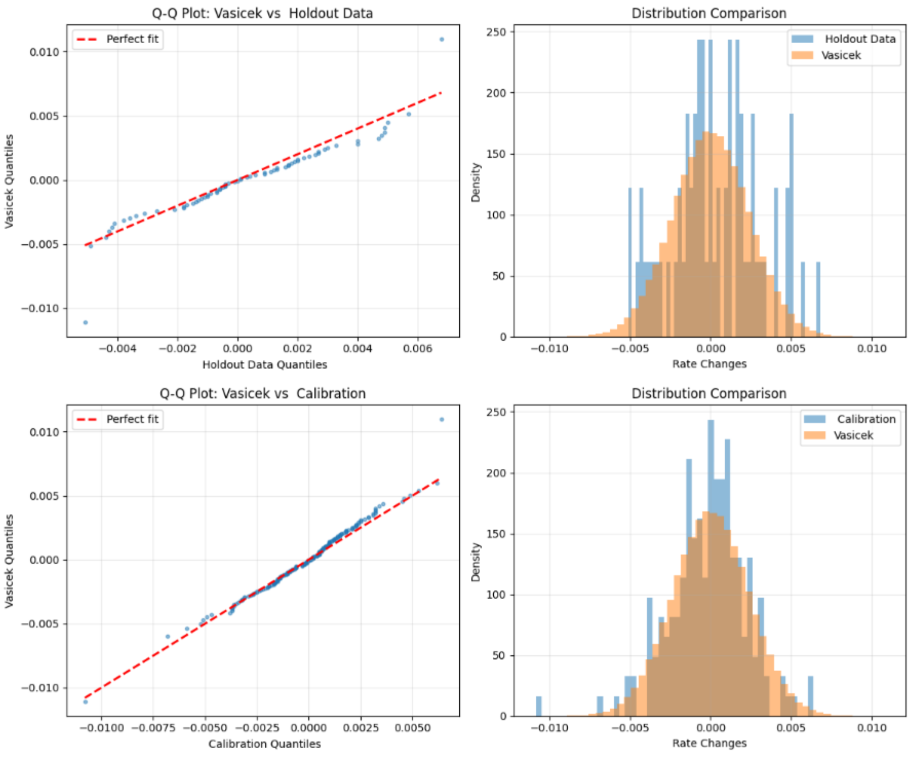
Output

Kolmonogorov-Smirnov Goodness of Fit Test

The Kolmonogorov-Smirnov (KS) test can be used to verify whether some sample distribution F_S, in our case the simulation, fits some theoretical distribution F_T, in our case the training, F_C, and test, F_H ,data. For the Vasicek model, F_V:

- Against F_C: $p = 0.49$ - Do not reject null hypothesis, insuffucient evidence to assume that the two distributions are not equal.
- Against F_H: $p = 0.49$ - Do not reject null hypothesis, insuffucient evidence to assume that the two distributions are not equal.

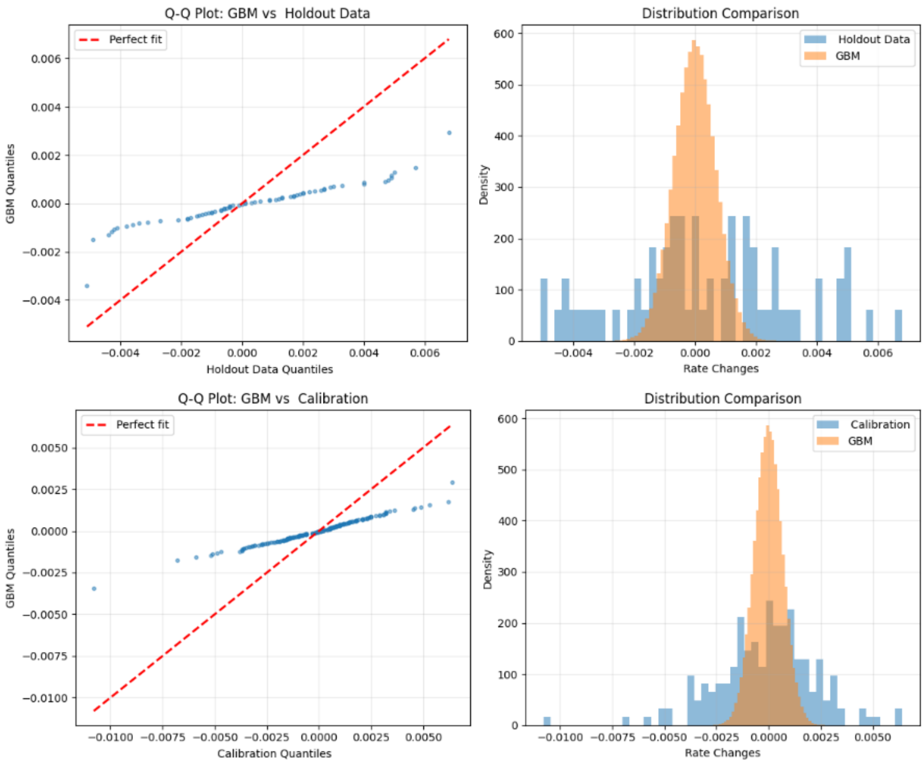
Interestingly, despite the failure of the training data set on the stationarity test, the stationary Vasiceek model passes the Kolmonogrov-Smirnov test against both the training and test data. The Q-Q plots are:



Testing our challenger GBM model, F_G, we find:

- Against F_C: $p = 0$ - Do not accept null hypothesis, suffucient evidence to assume that the two distributions are not equal.
- Against F_H: $p = 0$ - Do not accept null hypothesis, suffucient evidence to assume that the two distributions are not equal.

The Geometric Brownian Motion model, chosen to capture the non-stationarity of the rates fails epically in the KS test. Which can be even more apparently shown in the QQ plots:



4.1 Out of Sample Forecasting

4.2 Pensions

Implementation

The current implementation in this model is very primitive and not representative of realitlity.

- Assets
  - Modelled as only growing at the risk-free rate of inflation, whilst most portfolios are filled with much riskier assets
  - No contributions
- Liabilities
  - Treats liabilities with a fixed duration and convexity
  - No cash flow modelling

I advise that the following be added to this section of the model:

-

Parameter Sensitivity

I used the SensitivityAnalyzer in the sensitivty library in order to calculate the median funding ratios and percentage of severely underfunded portfolios (<80% funding ratio) in 225 different cases:

- Mean Reversion Speed: 0.18, 0.26, 0.34(base case), 0.42, 0.50
- Long Term Mean: 1.5%, 2%, 2.5%(base case), 3%, 3.5%
- Volatility: 0.4%, 0.6%, 0.8% (base case), 1%, 1.2%, 1.4%, 1.6%,1.8%, 2%

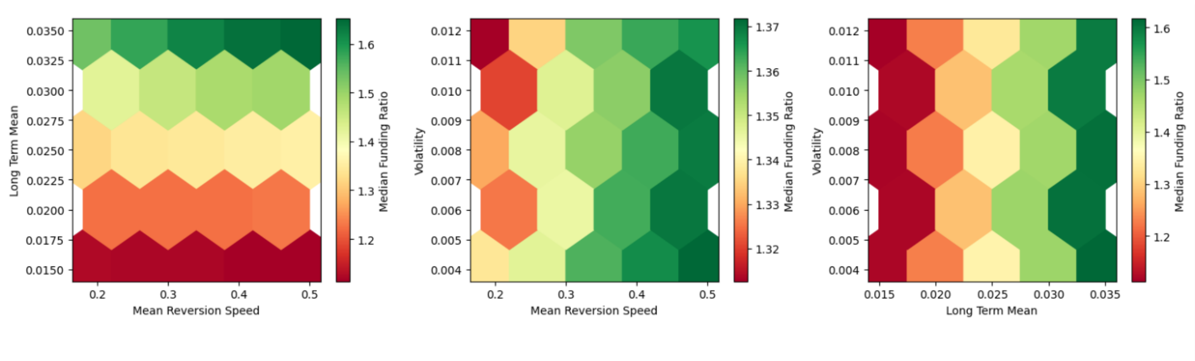


Figure 4.1: Median Sensitivity

As to be expected, the median funding ratios are most heavily affected by changes to the long-term mean.

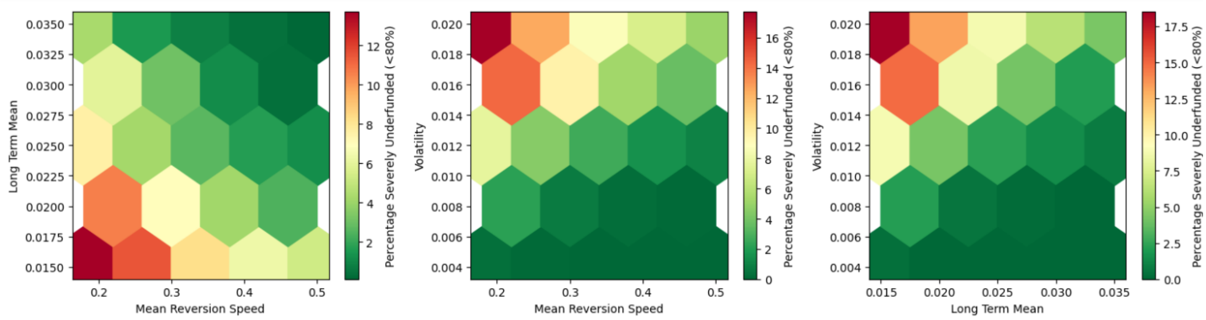


Figure 4.2: Underfunding Sensitivity

The percentage of portfolios severely underfunded is highly sensitive to the high volatility and low long term meaan regimes

Chapter 5

Documentation

Chapter 6

Governance

Chapter 7

this is my report