Akerman_George_00978849.pdf

by George Akerman

Submission date: 13-Sep-2022 03:13PM (UTC+0100)
Submission ID: 185943582
File name: Akerman_George_00978849.pdf (596.33K)
Word count: 18749
Character count: 88053

IMPERIAL COLLEGE LONDON

DEPARTMENT OF MATHEMATICS

Empirical Bond Pricing with Affine Models

Author: George Akerman (CID: 00978849)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2021-2022

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

With thanks to all who provided help and feedback during this thesis, in particular to my supervisors; Romano Trabalzini and Camilo Hernández.

Abstract

Forecasting bond prices is a key activity for fixed income traders. Although it is perhaps an ill posed question to ask what the true market value of a bond should be, based on the prevalent market conditions, it is possible to consider the relative value of bonds compared to other, similar maturity bonds on the same curve. This practice is referred to as relative value analysis. During relative value analysis, detected overpriced bonds are described as rich whereas under priced bonds are described as cheap. Common methods to conduct rich cheap analysis involve fitting a parsimonious parameterised curve to the observed empirical bond prices, and using this curve to infer what the value of each bond should be given the value of the other bonds. This process must be repeated multiple times on multiple different cross-sections of the yield curve. This is because certain bonds, due to issues around liquidity for instance, may be trading persistently rich or cheap. Therefore, we must consider the time series of rich cheap indicators, rather than just a single cross-section. A further issue with this approach is that when fitting the yield curve, only a single cross-section of yields is used. Because of this, the implied time series of yield curves is not self consistent; i.e., there is no way of ensuring this time series is arbitrage free in a multi period setting. You also fall into traps of comparing apples to oranges, in the sense that if a bond appears rich compared to where it has traded previously, you cannot detect weather this is due to the market having mispriced this bond, or a change in market conditions (regime change). A regime change would mean that the historic yield curves should have little to no bearing on current decision making.

To take steps to address this problem, we introduce affine models, as a way of detecting regime changes. Affine models can be fit to an entire time series of yield curves, not just a single snapshot, in such a way so as to be arbitrage free by construction. Further, the relationships required for the model, can be learnt by considering a window of training data. On the testing set, the output of the model can be directly compared to the more parsimonious snapshot models, allowing us to form an opinion on weather the snapshot models are adequately describing the true relationship. In cases where the difference between the snapshot models and the affine models is small, we can be more inclined to believe the results of our relative value analysis.

This thesis is organised as follows. In chapter 1, we define the basic notation and terminology around bonds and yield curve modelling that we will be using throughout. Chapter 2 is devoted to building intuition for the various problems faced by bond traders, and the different families of models that can be deployed to tackle each one. In chapter 3, we introduce the affine model framework, and look in detail at the ACM model specification and estimation procedure. We present the results of this estimation for South African local bonds in chapter 4 and in chapter 5, we attempt to interpret these results to give trading signals for rich cheap analysis. In chapter 6, we discuss some extensions that have been proposed in or inspired by the literature, that we would like to explore in the future. Finally, we give some concluding remarks in chapter 7.

Contents

1 Basics of bonds and rates **6**
1.1 Definitions . . . 6
1.1.1 Bonds and rates . . . 6
1.1.2 State price deflator . . . 7
1.2 Understanding the bond price . . . 11
1.2.1 Convexity . . . 11
1.2.2 Convexity and Itô . . . 11
1.3 What else affects yields . . . 12

2 Rationale of Yield curve modelling **13**
2.1 Yield curve modeling . . . 13
2.2 Taxonomy of yield curve models . . . 14
2.2.1 Taxonomy of yield curve models . . . 15
2.3 $\mathbb{P}$-measure vs $\mathbb{Q}$-measure . . . 16

3 Affine models for yield curve modelling and the ACM term structure model **17**
3.1 Affine models . . . 17
3.2 The ACM discrete time model for bond yields . . . 19
3.2.1 The model description . . . 19
3.2.2 Model estimation . . . 21
3.2.3 Term premium representation . . . 23

4 ACM output and fitting error **24**
4.1 Data . . . 24
4.2 Principal components . . . 24
4.3 State variables . . . 24
4.4 Goodness of fit . . . 25

5 Rich-cheap analysis and the ACM model **29**
5.1 Rich-cheap analysis . . . 29
5.2 Rich-cheap indicators . . . 29
5.3 Smoothing of rich cheap indicators . . . 31
5.4 Difficulties and criticism . . . 32
5.4.1 Emerging markets idiosyncrasies . . . 34
5.4.2 Numerical instability . . . 34
5.4.3 Avoidance of regime shifts . . . 34
5.4.4 Validation methodology . . . 35

6 Proposed extensions of the ACM model **37**
6.1 Non-spanning variables . . . 37
6.1.1 Theoretical challenges to ACM model extension . . . 37
6.1.2 Which factors to choose . . . 38
6.2 From data to state variables . . . 38
6.2.1 Interpretability . . . 38
6.2.2 Non-linearity, machine learning and autoencoders . . . 39

7 Conclusion 40

Bibliography 44

List of Figures

4.1 The input and output yield curves for the ACM model fit to ZAR bond data. The observed yields are calculated using a fitted Nelson-Siegel-Svensson model. The red line indicates the start of the out of sample period. 26
4.2 The fit quality for both in sample (left of red line) and out of sample (right of red line) for 2 separate fittings of the model to ZAR bond data at different times, compared to the time series of Nelson-Siegel-Svenson snapshot yield curves. The fit quality is the average of the squares of the difference between the input and output yield curves at each instance in time. Unsurprisingly, both plots show a clear spike around the time of the lock-downs caused by the Coronavirus pandemic. 27
4.3 The average fit quality for all out of sample cross-sections for ZAR bond data, the time on the x axis is the final date of the out of sample period. Interestingly, the fit quality clearly starts to deteriorate around February to March 2022. This deterioration can likely be attributed to a gradually increasing proportion of the out of sample data being from the period of time following the Russian invasion of Ukraine, whereas the model was fit predominantly during pre-invasion market conditions. The spike during June 2022 highlights a key problem, that the model is numerically unstable. If the training period does not contain enough data, resulting in the wrong relationship being learnt by the various regressions, then the implied bond yields can diverge from the snapshot model's yield curves very quickly. 28

5.1 The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for two South African bonds over time. 30
5.2 The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for a selection of bonds. The salient feature to observe is that the bottom four bonds in the figure, appear to have roughly constant rich cheap indicators up to 2022-07-18 whereafter they all simultaneously become more rich. This conclusion is not easy to swallow; therefore, we hope when we validate the rich cheap indicators, this latter portion will not pass the validation procedure. 32
5.3 The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for a selection of bonds, after the time series has been validated by taking only the observations for days with fitting errors in the lowest 50%. We notice that the period with the unrealistic simultaneous uptick in rich cheap indicator, did not pass the validation procedure. As a result, the time series now runs only to 2022-07-18 in contrast to figure 5.2 that had data all the way to 2022-08-12. 33
5.4 The average fit quality for all out of sample cross-sections for ZAR bond data, when the training window was reduced to 18 months only. The time on the x axis is the final date of the out of sample period. Notice that the fit error axis is now logarithmic. When the fit quality is this bad, the model is clearly unusable for any practical purpose. 35

List of Tables

5.1 A table showing the results of rich cheap analysis conducted on 2022-08-12. 31
5.2 A table showing the results of rich cheap analysis conducted on 2022-07-18. . . . 31
5.3 A table showing the result of rich cheap analysis, with validation procedure, conducted on 2022-07-18. 33

Chapter 1

Basics of bonds and rates

In this chapter, we will give some definitions and notation that shall be used throughout. We will also enter into a discussion about what factors effect the level and shape of the market observed yield curve. Later, this will allow us to decompose (3.2.9), the equation for excess returns in the ACM model, into contributions from the various factors.

1.1 Definitions

1.1.1 Bonds and rates

Definition 1.1.1. We denote by $P_t^{(\tau)}$ the price at time t of a zero-coupon bond that pays the holder a single payment of 1 unit of currency at maturity time $T := t + \tau$.

Consider for a second that we are living in a default free world, and we execute the following simple trading strategy. At time t, we invest an amount $P_t^{(\tau+\Delta T)}$ in a zero-coupon bond with maturity $T + \Delta T$ and finance this purchase by going short the same amount in zero-coupon bonds with maturity T. Then at time T, our short position matures so we close it by paying

$$u := \frac{P_t^{(\tau+\Delta T)}}{P_t^{(\tau)}}$$

before finally at time $T + \Delta T$, when our purchased bond matures, we receive a payment of 1. Using this strategy we guarantee ourselves a risk-free return of $1/u$. Let us define this rate of return as the continuously compounded forward rate $x_T^{T+\Delta T}$, so that

$$\exp\left(x_T^{T+\Delta T} \Delta T\right) = \frac{1}{u} = \frac{P_t^{(\tau)}}{P_t^{(\tau+\Delta T)}}$$

and

$$x_T^{T+\Delta T} = \frac{\ln P_t^{(\tau)} - \ln P_t^{(\tau+\Delta T)}}{\Delta T}.$$

Finally, we can define the instantaneous forward rate as the limit of $x_T^{T+\Delta T}$ as $\Delta T \to 0$.

Definition 1.1.2. [62, p.32] The time-t continuously compounded instantaneous forward rate for expiry T is given by

$$f_t^T := -\frac{\partial \ln P_t^{(\tau)}}{\partial T}.$$

Definition 1.1.3. The time-t instantaneous short rate is defined by

$$r_t = \lim_{T \to t} f_t^T.$$

The short rate is sometimes referred to as the risk free rate. It is the theoretical rate of return that an investor can receive, without any exposure to risk.

Next, we wish to derive a formula for the price of a very short dated zero-coupon bond in terms of the short rate. In a discrete time model, we can consider the function $\tau \to P_t^{(\tau)}$ to be defined by a linear interpolation through the points $\{P_t^{(\tau)}\}_{\tau\in\mathbb{Z}}$ so that

$$\begin{aligned} r_t &= \lim_{T\to t} f_t^T \\ &= \lim_{\tau\to 0}\lim_{\Delta T\to 0} \frac{\ln P_t^{(\tau)} - \ln P_t^{(\tau+\Delta T)}}{\Delta T} \\ &= \lim_{\tau\to 0}\left(\ln P_t^{(\tau)} - \ln P_t^{(\tau+1)}\right) \\ &= \ln P_t^{(0)} - \ln P_t^{(1)} \\ &= -\ln P_t^{(1)} \end{aligned}$$

because $P_t^{(0)} = 1$. A similar line of reasoning also works in continuous time. For an arbitrarily small time step Δt we have that

$$r_t = -\frac{1}{\Delta t}\ln P_t^{(\Delta t)} \qquad (1.1.1)$$

which implies

$$P_t^{(\Delta t)} = \exp\left(-r_t \Delta t\right). \qquad (1.1.2)$$

1.1.2 State price deflator

In this section we derive an equation for the pricing kernel M_t such that

$$P_t^{(\tau)} = \mathbb{E}\left[M_{t+\Delta t} P_{t+\Delta t}^{(\tau-\Delta t)}\right] \quad \forall\ \Delta t \in [0,\tau]. \qquad (1.1.3)$$

By doing so, we will gain a better understanding of how the market prices exposure to risks, and therefore we get one step closer to an understanding of what contributes to the market price of bonds. What follows is an adaptation of the approach given in [62, 13.3-13.9]. To achieve this objective, we will use Arrow-Debreu securities [19, section 9.2], which are theoretical securities paying 1 unit of currency in a particular state of the economy. As usual, we work in the filtered probability space given by

$$\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}^+}, \mathbb{P}\right)$$

where $\mathbb{R}^+$ is the set of all positive real numbers including 0. We make the assumption that Ω is finite so that expectations over Ω can be calculated as a sum over a finite number of atoms [49, p.224]. For convenience, let us assume that we are currently at time $t = 0$, although this assumption is not necessary. For any atom s, we denote by sp_s^T the price of an Arrow-Debreu security that pays one unit of currency at time T if the world at time T is in state s. Different subsets of Ω may be atoms for certain probability spaces but not others; however, our notation makes no reference to this as the particular sigma field we are considering when we take an expectation will be clear from context. As an aside, this assumption that Ω is finite is not actually restrictive, as if we assume that the security is measurable then we can represent it as the limit of securities taking only a finite number of different values, all the results we state here shall hold in the limit as well. If we consider an arbitrary security that pays out path dependent amounts only at time T with price v_0, it is clear by replication that

$$v_0 = \sum_{s\in\Omega} \text{payoff}_s sp_s^T \qquad (1.1.4)$$

where payoff_s is the payoff of our security at time T. Let p_s be the probability of ending up in state s and let us define

$$\pi_s^T = \frac{sp_s^T}{p_s^T}. \qquad (1.1.5)$$

The interpretation of π_s^T is as the value given by investors at time 0 to the receipt at time τ of one unit of currency in state s. Typically, we have $\pi < 1$ as investors usually prefer to receive their money now, rather than wait to receive it in the future. We call π the state price deflator because it 'deflates' the payoff of any security to the value that investors will give it at a previous time. Again, we appeal to the replication argument to show that this interpretation makes π_s^T well defined. From (1.1.4) and (1.1.5) we now have

$$v_0 = \mathbb{E}_0^{\mathbb{P}}\left[\text{payoff}_s \pi_s^T\right] \tag{1.1.6}$$

where

$$\mathbb{E}_t\left[\,\cdot\,\right] = \mathbb{E}_t\left[\,\cdot\,|\mathcal{F}_t\right].$$

If we specify the arbitrary security we are considering to be the zero-coupon bond with maturity $T = t + \tau$, then (1.1.6) becomes

$$P_0^{(\tau)} = \mathbb{E}_0^{\mathbb{P}}\left[\pi_s^T\right] = \sum_{s\in\Omega} \pi_s^T p_s^T$$

More generally, we can introduce our dependence on the current time t to give

$$P_t^{(\tau)} = \mathbb{E}_t^{\mathbb{P}}\left[\pi_s^T(t)\right] = \sum_{s\in\Omega} \pi_s^T(t) p_s^T. \tag{1.1.7}$$

In a frictionless market (which is an overriding assumption to everything we do here), an investor does not care weather she holds at time t a zero-coupon bond with time to maturity τ, or a security that pays out at time $t+\Delta t$ an amount equal to $P_{t+\Delta t}^{(\tau-\Delta t)}$. Therefore, we can also write

$$P_t^{(\tau)} = \mathbb{E}_t^{\mathbb{P}}\left[\pi_s^{t+\Delta t}(t) P_{t+\Delta t}^{(\tau-\Delta t)}\right] \tag{1.1.8}$$

and after comparing with (1.1.3) we identify

$$M_{t+\Delta t} = \pi_s^{t+\Delta t}(t).$$

Finally, we introduce the following identity that will be useful in the next section. If we apply (1.1.6) to a security that pays 1 currency at time T, only if some event A happens, then its price will be given by

$$x_t = \mathbb{E}_t^{\mathbb{P}}\left[\mathbb{1}_A \pi_s^T(t)\right].$$

But we also have that

$$x_t = \mathbb{E}_t^{\mathbb{P}}\left[x_{t+\Delta t}\pi_s^{t+\Delta t}(t)\right]$$

therefore $\forall A \in \mathcal{F}_T$,

$$\begin{aligned}
\mathbb{E}_t^{\mathbb{P}}\left[\mathbb{1}_A \pi_s^T(t)\right] &= \mathbb{E}_t^{\mathbb{P}}\left[x_{t+\Delta t}\pi_s^{t+\Delta t}(t)\right] \\
&= \mathbb{E}_t^{\mathbb{P}}\left[\mathbb{E}_{t+\Delta t}^{\mathbb{P}}\left[\mathbb{1}_A \pi_s^T(t+\Delta t)\right]\pi_s^{t+\Delta t}(t)\right] \\
&= \mathbb{E}_t^{\mathbb{P}}\left[\mathbb{1}_A \pi_s^T(t+\Delta t)\pi_s^{t+\Delta t}(t)\right].
\end{aligned}$$

Now, assuming that π_s^t is $\mathcal{F}_t$-measurable, we have that

$$\pi_s^T(t) = \pi_s^T(t+\Delta t)\pi_s^{t+\Delta t}(t) \quad \mathbb{P}-\text{a.s.} \quad \forall t, \Delta t \tag{1.1.9}$$

using [67, corollary A.1.5.3].

Solving for the state price deflator

Given that the State price deflator must be a positive value (otherwise investors are assigning a negative value to future money which would be a clear arbitrage), we hypothesise that π_s^t, follows a multivariate geometric Brownian motion [63, section 10.3.2]

$$\frac{d\pi_s^t(0)}{\pi_s^t(0)} = \mu_\pi(t)dt + \sigma_\pi(t)' dZ_t \tag{1.1.10}$$

where both $\mu_\pi(\tau)$ and $\sigma_\pi(\tau)$ can also have an additional dependency on π_s^t. Again to simplify notation we have fixed the current time to be time 0, but this is not necessary. Our first task is to obtain expressions for μ_π and σ_π. From (1.1.7) and (1.1.9), we have that for an arbitrarily small time step Δt,

$$P_t^{(\Delta t)} = \mathbb{E}_t^{\mathbb{P}}\left[\pi_s^{t+\Delta t}(t)\right] = \mathbb{E}_t^{\mathbb{P}}\left[\frac{\pi_s^{t+\Delta t}(t)\pi_s^t(0)}{\pi_s^t(0)}\right] = \mathbb{E}_t^{\mathbb{P}}\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}\right] \tag{1.1.11}$$

which implies, using (1.1.2) in the penultimate equality, that

$$\mathbb{E}_t^{\mathbb{P}}\left[\pi_s^{t+\Delta t}(0)\right] = \pi_s^t(0)P_t^{(\Delta t)} = \pi_s^t(0)\exp(-r_t\Delta t) = \pi_s^t(0)(1-r_t\Delta t) \tag{1.1.12}$$

where the last equality holds in the limit as $\Delta t \to 0$. Now, we can rewrite (1.1.10) as

$$\pi_s^{t+\Delta t}(0) - \pi_s^t(0) = \int_t^{t+\Delta t} 1\, d\pi_s^{t'}(0) = \int_t^{t+\Delta t} \pi_s^{t'}(0)\mu_\pi(t')\,dt' + \int_t^{t+\Delta t} \pi_s^{t'}(0)\sigma_\pi(t')'\,dZ_{t'}$$

where t' has been introduced only as a variable of integration. The integrals above can be discretised by assuming that over a sufficiently small time step, the integrands are constant. Therefore, we get that

$$\begin{aligned} \pi_s^{t+\Delta t}(0) - \pi_s^t(0) &\approx \pi_s^t(0)\mu_\pi(t)\Delta t + \pi_s^t(0)\sigma_\pi(t)'\left(Z_{t+\Delta t} - Z_t\right) \\ \implies \qquad \pi_s^{t+\Delta t} &\approx \pi_s^t(0)\left[1 + \mu_\pi(t)\Delta t + \sigma_\pi(t)'\left(Z_{t+\Delta t} - Z_t\right)\right]. \end{aligned}$$

In the limit as Δt converges to 0 from above, the approximation in the final step becomes exact. Therefore, we replace the approximation with an equality to obtain

$$\pi_s^{t+\Delta t}(0) = \pi_s^t(0)\left(1 + \mu_\pi(t)\Delta t + \sigma_\pi(t)'\left(Z_{t+\Delta t} - Z_t\right)\right) \tag{1.1.13}$$

and by taking expectations we get

$$E_t^{\mathbb{P}}\left[\pi_s^{t+\Delta t}(0)\right] = \pi_s^t(0)\left(1 + \mu_\pi(t)\Delta t\right)$$

because $\pi_s^t(0)$ is known at time t, meaning that we can take it outside the expectation. By equating this expression with (1.1.12) we see that

$$\mu_\pi(t) = -r_t. \tag{1.1.14}$$

Now from (1.1.8) and (1.1.9) we have

$$P_t^{(\tau)} = \mathbb{E}_t^{\mathbb{P}}\left[\pi_s^{t+\Delta t}(t)P_{t+\Delta t}^{(\tau-\Delta t)}\right] = \mathbb{E}_t^{\mathbb{P}}\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}P_{t+\Delta t}^{(\tau-\Delta t)}\right].$$

Therefore we get that

$$\begin{aligned} 1 &= \mathbb{E}_t\left[\frac{\pi_s^{t+\Delta t}(0)P_{t+\Delta t}^{(\tau-\Delta t)}}{\pi_s^t(0)P_t^{(\tau)}}\right] \\ &= \mathbb{E}_t\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}\right]\mathbb{E}_t\left[\frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] + \mathrm{Cov}\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}, \frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] \\ &= (1 - r_t\Delta t)\,\mathbb{E}_t\left[\frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] + \mathrm{Cov}\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}, \frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] \end{aligned} \tag{1.1.15}$$

using (1.1.12) in the last line. We will now deal with the remaining two expectation terms.

Similarly to above, we can also assume that

$$\frac{dP_t^{(T-t)}}{P_t^{(T-t)}} = \mu_P(t)dt + \sigma_P(t)'dZ_t$$

so that dicretisation gives

$$\frac{P_{t+\Delta t}^{(T-t-\Delta t)}}{P_t^{(T-t)}} = 1 + \mu_P(t)\Delta t + \sigma_P(t)'\left(Z_{t+\Delta t} - Z_t\right) \tag{1.1.16}$$

and taking expectations obtains

$$\mathbb{E}_t^{\mathbb{P}}\left[\frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] = 1 + \mu_P(t)\Delta t.$$

Finally, using (1.1.13) and (1.1.16) in the first line, we have

$$\begin{aligned}
\mathrm{Cov}\left[\frac{\pi_s^{t+\Delta t}(0)}{\pi_s^t(0)}, \frac{P_{t+\Delta t}^{(\tau-\Delta t)}}{P_t^{(\tau)}}\right] &= \mathrm{Cov}\left[1 + \mu_\pi(t)\Delta t + \sigma_\pi(t)'\left(Z_{t+\Delta t} - Z_t\right), 1 + \mu_P(t)\Delta t + \sigma_P(t)'\left(Z_{t+\Delta t} - Z_t\right)\right] \\
&= \mathrm{Cov}\left[\sigma_\pi(t)'\left(Z_{t+\Delta t} - Z_t\right), \sigma_P(t)'\left(Z_{t+\Delta t} - Z_t\right)\right] \\
&= \sigma_\pi(t)'\mathrm{I_d}\,\Delta t\,\sigma_P(t) \\
&= \sigma_\pi(t)'\sigma_P(t)\Delta t.
\end{aligned}$$

Equation (1.1.15) now becomes

$$1 = \left(1 - r_t\Delta t\right)\left(1 + \mu_P(t)\Delta t\right) + \sigma_\pi(t)'\sigma_P(t)\Delta t$$

which implies that

$$r_t = \mu_P(t) + \sigma_\pi(t)'\sigma_P(t). \tag{1.1.17}$$

In general, if Z contains k components, and thus so does σ_π and σ_P, then to solve for σ_π we would require k copies of the above equation. However, we can solve this equation in one dimension in which case

$$\sigma_\pi(t) = -\frac{\mu_P(t) - r_t}{\sigma_P(t)} = -\lambda_t \tag{1.1.18}$$

where λ_t is the Sharpe ratio [58, p.3]. This justifies why the Sharpe ratio is sometimes called the market price of risk [62, p.199], because the price of a bond depends only on the excess return and the volatility of the bond price, which is a proxy for the risk level. In general, large Sharpe ratios are unobtainable, meaning that high excess returns require high risk levels in order for λ_t to remain always within a reasonable upper bound. However, this interpretation only makes sense in a single factor economy where one doesn't have to specify which risk one is referring to. In a multi factor economy, different risks may have different prices in the market, that could even have different signs [35].

For now though we can solve (1.1.10). We denote by s_0 the single state that the economy is in at time 0 which is of course known at time 0. We have that

$$\begin{aligned}
\pi_s^t &= \pi_{s_0}^0 \exp\left[\int_0^t -r_u - \frac{1}{2}\sigma_\pi(u)'\sigma_\pi(u)\,du + \sigma_\pi(t)'Z_t\right] \\
&= \exp\left[\int_0^t -r_u - \frac{1}{2}\sigma_\pi(u)'\sigma_\pi(u)\,du + \sigma_\pi(t)'Z_t\right]
\end{aligned} \tag{1.1.19}$$

because $\pi_{s_0}^0 = 1$ An application of Itô's formula in multiple dimensions [42, p.153] will confirm this result is correct. Again we see that the price of a bond depends only on the excess returns and σ_π, each of whose parameters is a proxy to price of risk, for one of the underlying risk factors. The important observation is that (1.1.6) can be used to price any security not just bonds; therefore, differing prices between securities must have an interpretation in terms of different exposures to underlying risk factors. All underlying risks are priced in, and through (1.1.3) we can see that the only thing that causes different prices, is different levels of exposure to the various risk factors.

1.2 Understanding the bond price

This section is devoted to discussing the factors affecting the bond price. In the risk neutral world ($\mathbb{Q}$-measure), there is only one thing affecting bond prices and that is their future expectations. In the $\mathbb{Q}$ world, it is not possible to make returns in excess of the risk free rate, meaning that it is a trivial question which portfolio we should invest in to maximise returns because they are all equivalent. However, we do not live in a risk neutral world, and under the real world $\mathbb{P}$-measure things are different. Investors are in fact risk averse, meaning that they require a compensation (typically positive) for taking on any risk. Therefore, the yield realised on a bond portfolio is typically higher than the risk free rate, and the results we just derived for the state price deflator explain the effect on the realised yield of the exposure to different risk factors. There is however more to the story as the following derivation will highlight.

1.2.1 Convexity

The yield to maturity y_t^T for a zero-coupon bond is defined by [8, p.47]

$$P_t^T = \exp\left[-y_t^T(T-t)\right] \tag{1.2.1}$$

where the superscript T on the left hand side without brackets denotes maturity time, rather than time to maturity. It is natural to expect that the yield to maturity serves as a good approximation to the bond yield over a small time step. However, this is not the case as we shall now show.

Once again following [62, section 9.2], if we assume that our economy is subject to Brownian shocks of some description then the yield to maturity follows the following stochastic differential equation

$$dy_t^T = \mu dt + \sigma dZ_t$$

where μ and σ can both depend on t and y_t^T. Using Itô's formula we now obtain that

$$dP_t^T = y_t^T P_t^T dt - (T-t)P_t^T dy_t^T + \frac{1}{2}(T-t)^2 P_t^T \left(dy_t^T\right)^2$$

which implies

$$\frac{\mathbb{E}\left[dP_t^T\right]}{P_t^T} = \left(y_t^T - (T-t)\mu + \frac{1}{2}(T-t)^2\sigma^2\right) dt.$$

This shows that the expected yield on a bond depends, as expected, on the yield to maturity and the future expectations for the yield to maturity, but there is also a third term depending on the time to maturity and the volatility of the yield to maturity. This term is the convexity adjustment where

$$\text{convexity} = \frac{1}{2}\left(T-t\right)^2. \tag{1.2.2}$$

We also notice that the convexity adjustment is quadratic in the time to maturity, meaning for long dated bonds it becomes very significant, and to ignore its effects could be disastrous.

We have now seen the three building blocks for the bond yield: the expectation of future yields, the term premia (i.e. the priced exposure to risk factors) and the convexity adjustment.

1.2.2 Convexity and Itô

In this section, we will try to provide some intuition for convexity, as it is not immediately obvious where this effect comes from. As we see from the definition in (1.2.2), it results from the Itô correction term in Itô's formula and depends on the volatility of the yield to maturity. But if we recall equation (1.1.18), the risk premia also depended on the volatility of the risk factors? Is convexity just another priced risk factor or is it actually a different effect? The first thing to notice is that if the economy was generated, not by a Brownian motion, but by a process whose sample paths were almost surely of finite variation then the quadratic variation would be zero and the convexity term would have no effect. Nevertheless, there would still be a premium demanded for exposure to this risk factor as we would not be in a risk neutral world. This alone should be enough to convince us that convexity and risk premia are not manifestations of the same phenomena. Consider again our version of the economy where the risk factor y_t^T has 0 quadratic variation. What this means is that by observing the path of the risk factor at a finer and finer partition of points, our picture of the path gets clearer and clearer, and using linear interpolation between the observed points as an approximation for the path becomes progressively more accurate. For a fine enough partition, that linear interpretation accounts for all but an arbitrarily small amount of the total variation of the risk factor. Therefore, it is accurate enough to use the initial direction of the risk factor at t as a proxy for the direction of the yield factor change until $t + \Delta t$. In other words, we have that

$$dy_{t'}^T = dy_t^T \quad \forall t' \in [t, t + \Delta t) \, .$$

As a result it is fair to use a linear approximation for the effect of the risk factor on the bond price i.e.

$$dP_t^{(n)} = \frac{\partial P_t^{(n)}}{\partial t} dt + \frac{\partial P_t^{(n)}}{\partial y_t^T} dy_t^T .$$

However, when we allow the risk factor to have non-zero quadratic variation, the linear interpolation is no longer a sufficient approximation; because, no matter how fine we take our partition, the risk factor could still have infinite variation between the observed points. The Itô correction term, and hence convexity, arises as a correction for the inaccuracy of the linear interpolation technique.

The next question that it is natural to ask is what happens if this second order term still does not provide a significant correction? Well, thanks to Itô, we know that this could only occur if the quadratic variation of our risk factor was infinite. This in turn implies that the variance of the risk factor would go to infinity after any positive time interval. As much as this property would be undesirable from a modelling point of view it is unclear to what extent it is rejected by the empirical evidence [30], [64]. Nevertheless, the mathematical challenges of allowing such infinite variance are unappealing; therefore, the assumption that the economy is driven by Brownian shocks is commonplace.

1.3 What else affects yields

Our decomposition of yields described above is not necessarily complete. In this section, we will describe some of the effects we are choosing to ignore with our approach.

The first and perhaps most significant effect is liquidity premia. As part of our modelling assumptions, we assume that the market is frictionless [56]; however, this assumption can be hard to swallow, particularly in emerging markets. For example, on the run bonds trade at a slight premium to their off the run counterparts with the same maturities [71]. Investors, it would seem, require compensation for this liquidity risk [29]. For a model of the term structure that considers liquidity concerns see [18].

Another factor that may be omitted here is that of segmentation. Segmentation is the idea that supply and demand considerations can affect different parts of the yield curves at different times. For example, hedge funds and insurance firms may have particular preferred habitats on the yield curve and changes in the supply and demand within each habitat, caused by macroeconomic effects, may influence the shape of the yield curve. For an example of a model that considers these concerns see [69].

Chapter 2

Rationale of Yield curve modelling

2.1 Yield curve modeling

While zero-coupon bonds are traded instruments, the Zero-Coupon yield curve is typically constructed by interpolating market prices of the most liquidly traded instruments at different maturities. An accurate picture of the current cross-section of the yield curve undoubtedly provides important information for investors; however, how to correctly interpret the different common curve shapes is a matter of debate [50, Section 3.2].

Considering a default free world for a moment, if investors were certain of future yield curve shapes, then they would know, without uncertainty, for what price they would be able to liquidate any bond portfolio at any future time. It would therefore be a trivial optimisation problem to maximise returns over any investment horizon, the effects of no-arbitrage (ensuring sufficient liquidity[1]) would then ensure that all investment strategies would be essentially equivalent in the sense that it would be impossible to obtain extraordinary returns over any investment horizon. This is a particular application of the expectations theory of the term structure [8] which can be summarised as saying that current yields are no more than risk neutral expectations of future short rates. It is well known that the expectations theory of the term structure is not sufficient by itself to explain the yield curve as it assumes, among other things, that the price of risk is constant for all maturities. We know however, that longer duration bonds are more sensitive to changes in interest rates; therefore, the risk averse investor will require a higher compensation for the interest rate risk associated with a long term bond. This explains, at least partially, why the term structure of interest rates is more often than not upward sloping, and can be upward sloping even when the expectation is for future interest rates to stay the same or even decrease. Unfortunately, the risk premia for fixed maturities is in fact time varying [48] which further complexifys the task of forecasting future yields.

A complete and accurate understanding of the yield curve dynamics is also necessary for pricing portfolios of bonds, which is clearly vital for estimating holding period returns, but is also important for risk management and issues relating to regulatory compliance. Standard approaches often consider just the current cross-section of the time series, for instance, the regularisation approach introduced by Nelson and Siegel [55]. However, this approach fails to take into account earlier observations of the yield curve from which the current yield curve can be partially predicted [20]. Therefore, a more holistic approach that leverages this predictability should deliver more accurate risk metrics, as well as aid in the detection of bonds that are trading rich or cheap with respect to their fair price.

It is clear therefore that to understand this term structure properly we require a model that captures not just a cross-sectional snapshot of the yield curve, but explains the changes in rates over time. In addition, the ideal model would permit some sort of decomposition of the rates into a risk neutral component and a component compensating investors for risks (e.g. exposure to interest rates changes and liquidity risks etc.). Further, the model should be arbitrage free in the sense that an investor who incurs no risk cannot make a profit, or more precisely, the investor cannot

[1] This is likely a fair assumption. As of August 2020 the total value of outstanding debt in the global bond market was \$128.3tn [39].

make positive excess returns above the risk free interest rate over any holding period τ, where in a default free world, the risk free rate can be taken to be the yield to maturity of a zero-coupon bond with maturity τ.

2.2 Taxonomy of yield curve models

The first task for anyone who wishes to model bond yields is to decide upon the model they intend to use, or at the very least, decide upon which family of models they are going to draw from. One of the frustrations faced at this stage is that no single model can capture all the features that would be desirable. As a result, model choice necessarily becomes a trade-off, where some features are desired more highly than others, and some properties may be considered so indispensable for the application at hand, that we may be willing to accommodate other very undesirable features. Some of the features our modeller may desire include:

1. Quality of fit

   It sounds obvious but it is important that our model fits the observable data. A bad fit of the model to the time series would be indicative that the model is ill-specified, and is not sufficiently rich so as to capture the shape of the term structure. For some purposes, it may be important that our model recovers the current yield curve exactly. For instance, if we are attempting to price new securities with a methodology consistent with the current market price of traded securities, then it is important that the latter securities' prices are perfectly recovered by our chosen model. Therefore, we may choose to use a model that fits the current term structure perfectly. For example we could use the extended versions of the Vasicek, CIR or Dothan models [9, p.96-100].

2. Tractability

   For some applications, high frequency trading for example, speed is a key requirement for success [43]. Therefore, it is no use to have a model that can be calibrated perfectly to the market if the only way to infer things from it is via a computationally intensive Monte-Carlo estimation procedure. Not to mention, in the absence of analytic formulas, the results you obtain will be estimates anyway, thus ruining the perfect fit you may have had in the beginning. In cases where tractability is a requirement, simple short rate models e.g. Vasicek [68] or CIR [10] may be preferred.

3. Interpretability

   It is important when using any model to recognise that we are in fact only using a model, and that the real world is not the same as what our sterile modelling assumptions would suggest. Also, in the presence of critical political events that affect markets, a good past performance of a model does not indicate good future performance. Therefore, it is desirable that the parameters of a model, once fitted, have intuitive interpretations in the eyes of a trader. If a parameter allows an interpretation of long term volatility level say[2], then a trader who notices this level moving towards an unrealistic value following a re-calibration can avoid the potential disaster of using an ill-specified model before it occurs. A model, such as the libor market model, can use hundreds of parameters to fit the volatility surface, making any intuitive interpretation of those parameters very difficult.

4. Decomposablility

   The ability to decompose the bond prices coming out of a model into their constituent parts, affords a better understanding of the models output. As discussed above, the yield curve is an accumulation of multiple factors, and it can aid investment decisions to know to what degree each of these factors is contributing to the implied rates. For example, if you believe that a bond is undervalued, you may wish to go long in this bond; however, it would be of interest to know if that bond is also exposed to a risk factor for which a negative risk

[2] for example in the Vasicek model, the variance in the short rate at time t is given by $\frac{\sigma^2}{2k}\left[1 - \exp\left(-2kt\right)\right]$ [9, p.59] which converges to a long term level of $\frac{\sigma^2}{2k}$.

premium is being paid. If we did not require this exposure we may look to hedge it away. The model we discuss later in this paper permits an estimation of the term premium that investors receive. Term premium, typically positive, is the manifestation of the risk aversion that is built into bond prices. It is the difference between the price of a bond, and the price that it would be if investors were risk neutral. Term premium acts as a bridge between the two different probability measures used in finance namely the $\mathbb{P}$-measure, which is the real world probability measure, and the $\mathbb{Q}$-measure, typically used for pricing, under which today's prices are pure expectations of future prices discounted at the risk free rate.

2.2.1 Taxonomy of yield curve models

In this section we describe some of the common categories that term-structure models fall into.

1. Statistical models

   Statistical models, for example VAR, ARMA and GARCH, are designed to fit the observed time series very well and provide good predictive power. The nature of rates however do not lend themselves well to statistical models due to their quasi unit-root nature, meaning that the time series is non-stationary, as is the time series of differences [62, section 7.2.2]. In the case of rates, it is often considered that the first 3 principal components account for more than 95% of the variance in the bond yields [45]. The issue when statistical models are concerned is that the first principal component (level) mean reverts very slowly [62, p.114-5], making estimation of the coefficient for mean reversion very difficult [15].

2. Snapshot models [62, p.9]

   Snapshot models are parameterised curves designed to fit the observed yield curve at a particular snapshot in time. The fundamental paradox of snapshot models is that they ignore all the time series information available prior to the current time instance we are fitting. Therefore, in order to use these models, the parameters have to be refit regularly. However, the model provides nothing in the way of explaining the time varying characteristics of these parameters. In other words, there is no consistency over time built into the model. These models are important however as they allow for the inference of bonds at every point on the curve based on a finite set of observed bonds traded in the market. This can be a necessary requirement for the fitting of some more complex models [2]. Examples of snapshot models include Nelson-Siegel [55] and Fisher, Nychka and Zervos [28].

3. Equilibrium models

   Equilibrium models are models where the dynamics of the underlying driving variables are specified and calibrated to obtain a good fit with the observed data, but the requirement for a perfect fit is not enforced. Instead, a preference is shown towards having an endogenously consistent time series of yield curves [27]. Similarly to statistical models, they are designed to capture the current yield curve and describe how it will evolve through time. The current yield curve however, is endogenous to the model, meaning that it is obtained as an output of the model. These considerations are achieved by using simple, high level assumptions as starting points for the dynamics of the factors. For example, the Vasicek model states that the short rate mean reverts to a constant long term rate [68]. The simplicity of these assumptions allow the modeller to enforce (sometimes with additional assumptions) that the implied yield curves be consistent with each other and imply the absence of arbitrage. Equilibrium models often provide value to modellers when good data is sparse, or when liquidity premia or some other market microstructure effect may be disproportionately affecting certain bonds on the curve (as can be the case in emerging markets) [27]. In these cases, it would be illogical for most practical purposes to enforce that the yield curve perfectly reproduces the curve observed by the market.

4. Arbitrage-free models [14, p.99]

   Arbitrage-free models are similar to equilibrium models, but they have been designed to match the observed term structure of interest rates exactly [27]; i.e., there is no arbitrage between the model implied and the observed market prices of traded securities. This exact fit required that the initial yield curve be exogenous to the model, meaning it is used as an input when fitting the model. Arbitrage-free models typically specify dynamics for a series of state variables including the short rate. (In the case of short rate models, the short rate is the only state variable.) Arbitrage-free models are used to hedge derivatives, where an exact fit to market prices is crucial. Equilibrium models can often be modified into arbitrage-free models by the introduction of time varying parameters, for example the Hull-White extensions of Vasicek and CIR models [72] and the extension of the Dothan model to the Black-Derman-Toy model [6]. Necessarily this goodness of fit comes with a trade-off. In the case of the Hull-White extension of the Vasicek model, allowing a perfect fit to the current yield curve results in the loss of much of the analytical tractability [9, p.80-1]. Early examples of models in this category include Ho and Lee [34] and the celebrated Heath Jarrow Morton framework [32] [33].

2.3 $\mathbb{P}$-measure vs $\mathbb{Q}$-measure

Very crudely, a distinction begins to appear in the use cases of these models. In some instances, when the task at hand relates to pricing or hedging, we are not attempting a judgement about how the real world will move. For example, when you hedge a call option, this depends on the delta of the stock, which does not depend at all on whether or not the stock is likely to increase or decrease in value. In this case, you will be working with the risk neutral $\mathbb{Q}$-measure, under which all tradeable assets are martingales and arbitrage free pries can be found by taking expectations of future payoffs. However, if you are a hedge fund or investor looking to maximise returns, it clearly does matter what our view of the world is. In this case, you would be working under the $\mathbb{P}$-measure, which is the physical probability measure for the world in which we are living. In both cases the techniques would be similar, however, the estimates for the dynamics that you will postulate will most likely be different. Neither would be more correct than the other, they would just have different purposes. As we mentioned previously, the term premium can be seen as a bridge between the two measures. It is the adjustment you would have to make to go between the price under the $\mathbb{Q}$-measure, and the price that would be obtained if you used the $\mathbb{P}$-measure as the risk neutral measure.

Chapter 3

Affine models for yield curve modelling and the ACM term structure model

We now proceed to discuss a particular class of equilibrium term structure models known as affine term structure models, and we prove an expression for the bond price in affine models. The price of many interest rate derivatives, for example forward rate agreements [9, p.11] and interest rate swaps [9, section 1.5], can be written in terms of the bond price at various maturities. Therefore, it is the analytic tractability of this expression for the bond price that makes affine models desirable.

3.1 Affine models

Definition 3.1.1. [13] An affine model is a specification of bond prices in terms of an N-dimensional vector of state variables X_t, such that:

1. the instantaneous short rate r_t is given by an affine transformation

$$r_t := r(X_t) := u + g' X_t$$

where $g \in \mathbb{R}^N$ is a vector weights;

2. in the real-world measure $\mathbb{P}$, the state variables evolve according to

$$dX_t = \mu^{\mathbb{P}}(X_t)\, dt + \sigma(X_t) dZ_t^{\mathbb{P}}$$

where Z_t is a standard N-dimensional Brownian motion under $\mathbb{P}$, σ is an N x N matrix and

$$\mu^{\mathbb{P}}(X_t) := a^{\mathbb{P}} + b^{\mathbb{P}} X_t$$

is an affine transformation of the state variables with $b^{\mathbb{P}}$ an N x N matrix;

3. the covariances between the components of the Brownian shocks are given by an affine transformation of the state variables i.e.,

$$[\sigma(X_t)\sigma(X_t)']_{ij} = \alpha_{ij} + (\beta_{ij})' X_t$$

where β_{ij} is an N-dimensional vector.

4. There exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the bond price is given by

$$P_t^{(\tau)} = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r(X_s)\, ds\right)\middle| X_t\right] \tag{3.1.1}$$

and item 2 holds with $\mathbb{P}$ replaced by $\mathbb{Q}$ throughout i.e.,

$$dX_t = \mu^{\mathbb{Q}}(X_t)\,dt + \sigma(X_t)dZ_t^{\mathbb{Q}}$$

and

$$\mu^{\mathbb{Q}}(X_t) := a^{\mathbb{Q}} + b^{\mathbb{Q}}X_t.$$

What follows is the key result regarding affine models, that makes their usage so widespread.

Theorem 3.1.2 ([24])**.** *In an affine model where X can take all values in some open set, there exists functions $A(\tau)$ and $B(\tau)$ such that*

$$P_t^{(\tau)} = \exp\left[A(\tau) + B(\tau)'X_t\right]. \tag{3.1.2}$$

Proof. The multidimensional Feynman-Kac theorem [59, p.25] says that $P_t^{(\tau)}$ defined in (3.1.1) is the solution to

$$\frac{\partial P_t^{(\tau)}}{\partial t} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sigma_i(X_t)'\sigma_j(X_t)\right)\frac{\partial^2 P_t^{(\tau)}}{\partial(X_t)_i\partial(X_t)_j} + \mu^{\mathbb{Q}\prime}\frac{\partial P_t^{(\tau)}}{\partial X_t} - r(X_t)P_t^{(\tau)} = 0. \tag{3.1.3}$$

Therefore, (3.1.2) holds if and only if (3.1.3) is satisfied by $P_t^{(\tau)}$ given by (3.1.2). In other words (3.1.2) holds, if and only if

$$\left(-A_\tau(\tau) - B_\tau(\tau)'X_t + B(\tau)'\mu^{\mathbb{Q}}(X_t) + \frac{1}{2}B(\tau)'\sigma(X_t)\sigma(X_t)'B(\tau) - r(X_t)\right)P_t^{(\tau)} = 0 \tag{3.1.4}$$

where A_τ and B_τ are the derivatives of A and B respectively. Therefore, the statement we wish to prove is equivalent to the statement that there exists A and B satisfying (3.1.4), which in turn is equivalent to

$$-A_\tau(\tau) - B_\tau(\tau)'X_t + B(\tau)'\mu^{\mathbb{Q}}(X_t) + \frac{1}{2}B(\tau)'\sigma(X_t)\sigma(X_t)'B(\tau) - r(X_t) = 0 \tag{3.1.5}$$

because $P_t^{(\tau)}$ is strictly positive. From our assumptions on μ, σ and r we can see that the LHS of (3.1.5) is affine in the state variables. We can therefore gather together separately the terms including/excluding X_t and set both parts to 0 because 3.1.5 must hold for all (X_t, τ) on some open set, and an affine function can only equal 0 on an open set if the function itself is identically 0 everywhere. We look therefore for solutions to

$$B_\tau(\tau) = \mathcal{B}(B(\tau)) \tag{3.1.6}$$

and

$$A_\tau(\tau) = \mathcal{A}(B(\tau)) \tag{3.1.7}$$

where all the terms of $\mathcal{B}(B(\tau))$ and $\mathcal{A}(B(\tau))$ are linear quadratic for example

$$\mathcal{B}(B(\tau))_i = a + \sum_{j=1}^{N}\left(b_j B_j(\tau)\right) + \sum_{j=1}^{N}\sum_{k=1}^{N}\left(d_{jk}B_j(\tau)B_k(\tau)\right).$$

To derive initial conditions for (3.1.6) and (3.1.7) we consider (3.1.2) with $\tau = 0$ to see that $A(0) = B(0) = 0$. Further, for any given B we can recover A as

$$A(\tau) = \int_0^\tau \mathcal{A}(B(s))\,ds.$$

Finally, we have reduced our problem to finding a solution to (3.1.6) which is known as a Riccati equation with constant coefficients. The solution to this class of equations is given in [53, p.242-3]. □

Once again, it is the analytical tractability of the expression given in theorem 3.1.2 that provides much of the value of affine models. As a consequence of theorem 3.1.2, pricing bonds in an affine model is equivalent to determining the parameters $A(\tau)$ and $B(\tau)$.

3.2 The ACM discrete time model for bond yields

In this section we will introduce the ACM affine term structure model. The ACM model dynamics are the same as are used by many authors e.g. [21] [41], however the differences appear in the estimation procedures. Remember that the task at hand is to determine the parameters $A(\tau)$ and $B(\tau)$, which will require some assumptions to be made along the way.

3.2.1 The model description

The following model is described in [2] and is commonly referred to as the ACM model. In this chapter we will highlight some of the key equations and expositions. It is important to keep in mind the purpose of this model, which is to take in a vector of yields, along with a historical time series of yields, and re-produce the zero-coupon yield curve time series. This time series may differ from the curves used as input, but will fit the input closely as well as be free from arbitrage opportunities. For more detail, please refer to the original paper [2].

Suppose that in discrete time, our N-dimensional time dependent vector of state variables denoted X_t evolves according to

$$X_{t+1} = \mu + \phi X_t + v_{t+1} \tag{3.2.1}$$

where

$$v_{t+1}|\mathcal{F}_t \sim \mathcal{N}(0, \Sigma)$$

and $\mathcal{F}_t$ is the filtration generated by the process $\{X_s\}_{s=0}^{t}$. We also assume that bond prices are given by

$$P_t^{(n)} = \mathbb{E}_t \left[M_{t+1} P_{t+1}^{(n-1)} \right] \tag{3.2.2}$$

where

$$M_{t+1} = \exp\left(-r_t - \frac{1}{2}\lambda_t' \lambda_t - \lambda_t' \Sigma^{-\frac{1}{2}} v_{t+1} \right) \tag{3.2.3}$$

consistent with (1.1.19), and

$$\lambda_t := \Sigma^{-\frac{1}{2}} \left(\lambda_0 + \lambda_1 X_t \right). \tag{3.2.4}$$

In other words, the market price of risk is affine in the state variables. In this section, the superscript (n) on the bond price $P_t^{(n)}$ once again denotes the time to maturity.

If we buy a long maturity bond and sell it one period later, the log return we will receive denoted $x_{t+1}^{(n-1)}$ is defined by

$$x_{t+1}^{(n-1)} := \ln \frac{P_{t+1}^{(n-1)}}{P_t^{(n)}} = \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)}.$$

We denote the excess log return of this strategy above the short rate as

$$rx_{t+1}^{(n-1)} := x_{t+1}^{(n-1)} - r_t = \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} - r_t. \tag{3.2.5}$$

Our first task is to obtain an exprseeion for the excess returns, in terms of the state variables, and parameters we will be able to estimate from market data. To that end, we use (3.2.5) and (3.2.3) in (3.2.2) to give

$$1 = \mathbb{E}\left[\exp\left(rx_{t+1}^{(n-1)} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma^{-\frac{1}{2}} v_{t+1} \right) \right] = \exp\left(-\frac{1}{2}\lambda_t'\lambda_t \right) \mathbb{E}\left[\exp\left(rx_{t+1}^{(n-1)} - \lambda_t'\Sigma^{-\frac{1}{2}} v_{t+1} \right) \right]$$

therefore

$$\exp\left(\frac{1}{2}\lambda_t'\lambda_t \right) = \mathbb{E}\left[\exp\left(rx_{t+1}^{(n-1)} - \lambda_t'\Sigma^{-\frac{1}{2}} v_{t+1} \right) \right] = \mathbb{E}\left[\exp\left(\begin{pmatrix} 1 & -\lambda_t'\Sigma^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} rx_{t+1}^{(n-1)} \\ v_{t+1} \end{pmatrix} \right) \right]$$

where the matrices on the right hand side are block matrices. It is common knowledge that financial excess returns are typically too heavy tailed to be accurately characterised by a normal distribution

[52, p.80]. Nevertheless we assume that $\left\{rx_{t+1}^{(n-1)}, v'_{t+1}\right\}$ are jointly normal so that an application of the characteristic function for normal distributions gives

$$\exp\left(\frac{1}{2}\lambda'_t\lambda_t\right) = \exp\left[\mathbb{E}\left[rx_{t+1}^{(n-1)}\right] + \frac{1}{2}\begin{pmatrix}1 & -\lambda'_t\Sigma^{-\frac{1}{2}}\end{pmatrix}\tilde{\Sigma}\begin{pmatrix}1 \\ -\Sigma^{-\frac{1}{2}}\lambda_t\end{pmatrix}\right] \tag{3.2.6}$$

where

$$\tilde{\Sigma} := \begin{pmatrix}\mathrm{Var}_t\left[rx_{t+1}^{(n-1)}\right] & \mathrm{Cov}_t\left[rx_{t+1}^{(n-1)}, v_{t+1}\right] \\ \mathrm{Cov}_t\left[rx_{t+1}^{(n-1)}, v_{t+1}\right]' & \Sigma\end{pmatrix}$$

is again a block matrix. Now we define

$$\beta_t^{(n-1)} := \Sigma^{-1}\mathrm{Cov}_t\left[rx_{t+1}^{(n-1)}, v_{t+1}\right]' \tag{3.2.7}$$

so that from (3.2.6) we have

$$\frac{1}{2}\lambda'_t\lambda_t = \mathbb{E}\left(rx_{t+1}^{(n-1)}\right) + \frac{1}{2}\left(\mathrm{Var}_t\left[rx_{t+1}^{(n-1)}\right] - \beta_t^{(n-1)\prime}\Sigma^{\frac{1}{2}}\lambda_t - \lambda'_t\Sigma^{\frac{1}{2}}\beta_t^{(n-1)} + \lambda'_t\lambda_t\right).$$

Now using (3.2.4) we obtain

$$\mathbb{E}_t\left(rx_{t+1}^{(n-1)}\right) = \beta_t^{(n-1)\prime}\left[\lambda_0 + \lambda_1 X_t\right] - \frac{1}{2}\mathrm{Var}_t\left[rx_{t+1}^{(n-1)}\right]. \tag{3.2.8}$$

From (3.2.7), we see that the conditional covariance of the excess returns with the return innovations is given by

$$\mathrm{Cov}\left[rx_{t+1}^{(n-1)} - \mathbb{E}_t\left(rx_{t+1}^{(n-1)}\right), v_{t+1}\middle|\mathcal{F}_t\right] = \beta_t^{(n-1)\prime}\Sigma$$

therefore we define

$$e_{t+1}^{(n-1)} := rx_{t+1}^{(n-1)} - \mathbb{E}_t\left(rx_{t+1}^{(n-1)}\right) - \beta_t^{(n-1)\prime}v_{t+1}$$

to see that

$$\mathrm{Cov}\left[e_{t+1}^{(n-1)}, v_{t+1}\middle|\mathcal{F}_t\right] = 0$$

and

$$rx_{t+1}^{(n-1)} - \mathbb{E}_t\left(rx_{t+1}^{(n-1)}\right) = \beta_t^{(n-1)\prime}v_{t+1} + e_{t+1}^{(n-1)}.$$

Next, we assume that for all t and n

$$\beta_t^{(n)} = \beta^{(n)},$$

meaning that the covariance of the excess returns with the pricing factor inovations, aswell as the variance of the excess returns themselves, are both constant in time. We also assume that the $e_t^{(n-1)}$ are conditionally (i.i.d) with variance σ^2 so that

$$\mathrm{Var}_t\left[rx_{t+1}^{(n-1)}\right] = \beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2.$$

Therefore, (3.2.8) becomes

$$rx_{t+1}^{(n-1)} = \beta^{(n-1)\prime}\left(\lambda_0 + \lambda_1 X_t\right) - \frac{1}{2}\left(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2\right) + \beta^{(n-1)\prime}v_{t+1} + e_{t+1}^{(n-1)}. \tag{3.2.9}$$

We have achieved the first goal we set for ourselves. In addition, the first three terms in (3.2.9) can be interpreted as the expected returns; an adjustment for convexity; and the priced exposure to risk factors respectively, while the last term is an error term. Recall that in chapter 1, we discussed how these three main components were the main contributors to the yield curve.

3.2.2 Model estimation

In order to estimate parameters based on all observations at times $t = 0, 1, \ldots, T$ simultaneously, we combine the observations of our state variables into a single matrix

$$X_- = (X_0|X_1|\ldots|X_{T-1}).$$

Using the operator $\mathrm{vec}(A)$ which transforms a matrix to a vector by concatenating the rows of the matrix. We define the matrix

$$rx := \beta' (\lambda_0 \mathbb{1}_T' + \lambda_1 X_-) - \frac{1}{2}\left(B^* \mathrm{vec}(\Sigma) + \sigma^2 \mathbb{1}_N\right)\mathbb{1}_T' + \beta' V + E$$

where $\mathbb{1}_Y$ is a Y-dimensional vector of 1s, $\beta = \left(\beta^{(1)}\middle|\beta^{(2)}\middle|\ldots\middle|\beta^{(N)}\right)$,

$$\begin{aligned} B^* &:= \left(\mathrm{vec}\left(\beta^{(1)}\beta^{(1)\prime}\right)\middle|\mathrm{vec}\left(\beta^{(2)}\beta^{(2)\prime}\right)\middle|\ldots\middle|\mathrm{vec}\left(\beta^{(N)}\beta^{(N)\prime}\right)\right)', \\ V &:= (v_1|v_2|\ldots|v_T), \end{aligned}$$

and

$$(E)_{ij} := e_j^{(i)}.$$

To justify the use of this notation, we see that componentwise

$$(rx)_{il} = (\beta'\lambda_0\mathbb{1}_T')_{il} + (\beta'\lambda_1 X_-)_{il} - \frac{1}{2}\left(B^* \mathrm{vec}(\Sigma)\mathbb{1}_T'\right)_{il} - \frac{1}{2}\left(\sigma^2\mathbb{1}_N\mathbb{1}_T'\right)_{il} + (\beta'V)_{il} + E_{il}$$

where we can simplify many of these terms considerably. We have

$$\begin{aligned} \left(B^* \mathrm{vec}(\Sigma)\mathbb{1}_T'\right)_{il} &= B^*_{ij}\,\mathrm{vec}(\Sigma)_{j1}\,(\mathbb{1}_T')_{1l} \\ &= B^*_{ij}\,\mathrm{vec}(\Sigma)_j \\ &= \mathrm{vec}\left(\beta^{(i)}\beta^{(i)\prime}\right)_j \mathrm{vec}(\Sigma)_j \\ &= \left(\beta^{(i)}\beta^{(i)\prime}\right)_{jk}\Sigma_{jk} \\ &= \beta_j^{(i)}\beta_k^{(i)}\Sigma_{jk} \\ &= \beta_j^{(i)}\Sigma_{jk}\beta_k^{(i)} \\ &= \beta^{(i)\prime}\Sigma\beta^{(i)}, \end{aligned}$$

$$\left(\sigma^2\mathbb{1}_N\mathbb{1}_T'\right)_{il} = \sigma^2$$

and

$$\begin{aligned} (\beta'\lambda_0\mathbb{1}_T')_{il} + (\beta'\lambda_1 X_-)_{il} + (\beta'V)_{il} + E_{il} &= (\beta')_{ij}(\lambda_0)_j(\mathbb{1}_T')_l + (\beta')_{ij}(\lambda_1)_{jk}(X_-)_{kl} + (\beta')_{ij}V_{jl} + e_l^{(i)} \\ &= \beta^{(i)\prime}\lambda_0 + \beta_j^{(i)}(\lambda_1)_{jk}(X_{l-1})_k + \beta_j^{(i)}(v_l)_j + e_l^{(i)} \\ &= \beta^{(i)\prime}\lambda_0 + \beta^{(i)\prime}\lambda_1 X_{l-1} + \beta^{(i)\prime}v_l + e_l^{(i)} \\ &= \beta^{(i)\prime}(\lambda_0 + \lambda_1 X_{l-1}) + \beta^{(i)\prime}v_l + e_l^{(i)}. \end{aligned}$$

Therefore,

$$(rx)_{il} = \beta^{(i)\prime}(\lambda_0 + \lambda_1 X_{l-1}) - \frac{1}{2}\left(\beta^{(i)\prime}\Sigma\beta^{(i)} + \sigma^2\right) + \beta^{(i)\prime}v_l + e_l^{(i)}$$

so that after comparing with (3.2.9)we see that

$$(rx)_{n-1,t+1} = rx_{t+1}^{(n-1)}. \tag{3.2.10}$$

In order to begin estimating our parameters, we use linear regression on the relationship shown in 3.2.1 to estimate $\hat{V}$, a least squares estimate for V. Further, we estimate Σ as

$$\hat{\Sigma} := \frac{\hat{V}\hat{V}'}{T}.$$

Next, we perform a linear regression using the relationship (3.2.10) to obtain parameter estimates $\hat{a}, \hat{\beta}$ and $\hat{c}$ of

$$a := \beta'\lambda_0 - \frac{1}{2}\left(B^* \operatorname{vec}(\Sigma) + \sigma^2 \mathbb{1}_N\right), \quad \beta \text{ and } c := \beta'\lambda_1$$

respectively, and we call $\hat{E}$ the residuals from this regression. We can then solve to obtain the estimates $\hat{\lambda}_0$ and $\hat{\lambda}_1$ by replacing parameters with their estimates where necessary and using

$$\hat{\sigma}^2 := \frac{1}{NT} \operatorname{Tr}\left(\hat{E}\hat{E}'\right).$$

Remember from our discussion at the end of section 1.1.2 that the parameter λ_t, in conjunction with (3.2.3), is all that is required to price any security whose value is known at a given time τ. We have now obtained estimates for all the components of λ_t given by (3.2.4) so we should now have everything we need to compute a yield curve, using Monte-Carlo simulation to sample the state variables. Nevertheless, we shall ignore equation (1.1.3) preferring instead a more direct approach. Our starting point will be (3.1.2) from theorem 3.1.2; however, as we are considering real data we cannot be sure that the affine assumptions all hold. To account for this we introduce an error term $u_t^{(n)}$ such that

$$\ln P_t^{(n)} = A(n) + B(n)'X_t + u_t^{(n)}. \tag{3.2.11}$$

This error term is often referred to as the measurement error [62, p.677] and reflects the difference between the real world price and the price implied by the model. Our task remains to obtain expressions for the functions A and B. We substitute (3.2.11) into (3.2.5) and use (1.1.1) to obtain

$$rx_{t+1}^{(n-1)} = A(n-1) + B(n-1)'X_{t+1} + u_{t+1}^{(n-1)} - A(n) - B(n)'X_t - u_t^{(n)} + A(1) + B(1)'X_t + u_t^{(1)}.$$

Then, after substituting in (3.2.1), we have

$$\begin{aligned} rx_{t+1}^{(n-1)} &= A(n-1) + B(n-1)'\left(\mu + \phi X_t + v_{t+1}\right) + u_{t+1}^{(n-1)} \\ &\qquad - A(n) - B(n)'X_t - u_t^{(n)} + A(1) + B(1)'X_t + u_t^{(1)} \\ &= \beta^{(n-1)\prime}\left(\lambda_0 + \lambda_1 X_t\right) - \frac{1}{2}\left(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2\right) + \beta^{(n-1)\prime}v_{t+1} + e_{t+1}^{(n-1)} \end{aligned}$$

where the last equality comes from (3.2.9). Rearranging then gives

$$\begin{bmatrix} A(n) \\ \\ +B(n)'X_t \\ +B(n-1)'v_{t+1} \\ +u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)} \end{bmatrix} = \begin{bmatrix} A(n-1) + B(n-1)'\mu - \beta^{(n-1)\prime}\lambda_0 \\ \qquad + \frac{1}{2}\left(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2\right) + A(1) \\ + \left(B(n-1)'\Phi + B(1)' - \beta^{(n-1)\prime}\lambda_1\right)X_t \\ +\beta^{(n-1)\prime}v_{t+1} \\ +e_{t+1}^{(n-1)} \end{bmatrix}.$$

Because t is arbitrary, as long as X takes values in some non-degenerate set[1] we can equate the above coefficients so that, after noticing in particular that $\beta^{(n)} = B(n)$,

$$A(n) = A(n-1) + B(n-1)'(\mu - \lambda_0) + \frac{1}{2}\left(B(n-1)'\Sigma B(n-1) + \sigma^2\right) + A(1) \tag{3.2.12}$$

$$B(n)' = B(n-1)'\left(\Phi - \lambda_1\right) + B(1)' \tag{3.2.13}$$

$$\beta^{(n)} = B(n)$$

[1]We do not need to assume this for v as we get this for free from the already assumed distribution for v.

and

$$u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)} = e_{t+1}^{(n-1)}.$$

These equations can be used to recursively determine $A(n)$ and $B(n)$ up to an arbitrarily large n in terms of $A(1)$ and $B(1)$ as well as parameters we have already determined. To calculate the initial values $A(1)$ and $B(1)$ we perform a regression based on (3.2.11) with $n = 1$.

practical advantages

One of the advantages of this estimation procedure, versus say a maximum likelihood estimator, is that at every stage there requires only Linear regressions or direct computations. At no point do we need to iteratively fit a function using a non-linear function solver, that would be computationally expensive as well as potentially introducing numerical instability [44]. This is an important consideration as it would allow for a model of this kind to be used as a live monitor for traders. Should there be a shift in the market, it is advantageous that traders can have up to date model output in seconds, rather than the hours or even days that it could take to re-calibrate other models, particularly if it involved the training of a deep neural network.

3.2.3 Term premium representation

So far we have been assuming that our investor is risk averse, and this risk aversion is incorporated in the the state price deflator (M_{t+1}) in (3.2.2), which acts as a 'bridge' between the real-world $\mathbb{P}$-measure and the risk-neutral $\mathbb{Q}$-measure. If we were to instead assume that these two measures coincided, then the only factor contributing to the bond price today, would be the future expectation of bond prices. These assumptions are equivalent to saying that M_{t+1} in (3.2.2) contains only the discounting by the risk free rate i.e.,

$$M_{t+1} = \exp(-r_t) \quad \Longrightarrow \quad \lambda_t = 0 \quad \Longrightarrow \quad \lambda_0 = \lambda_1 = 0.$$

Using the above, we can rerun the recursive equations (3.2.12) and (3.2.13) to obtain an estimate for the risk neutral yield curve. The term premium implied by the model is then given by the difference between these two yield curves.

Chapter 4

ACM output and fitting error

4.1 Data

In this section, we apply the model described in section 3.2 to the bond markets for South African Rand (ZAR) local bonds, which means they are bonds purchased in local currency that also pay their cash flows in that same local currency. We use data at the close of each day from 11th May 2018 - 12th August 2022 provided by Deutsche Bank via Bloomberg. We use a rolling window scheme with a window size of 2 years and 30 weeks, which is further divided into a two year training set and a 30 week out of sample testing set. We advance the window forward one day at a time, allowing a separate model to be trained and evaluated to give predictions for each of the 179 most recent business days. In each case, only information that would have been available to the investor at the time is used to make said predictions. We will discuss possible extensions of the input data in section 6.1 when we look at non-spanning factors. However, initially we will use just information contained in the yield curve, from which there is strong evidence for excess bond holding return predictability [25] [12] [61].

4.2 Principal components

In order to apply the methodology described in section 3.2, we need to transform our yields into a set of state variables. Ideally, we would also like to reduce the dimension of the input to avoid over fitting. Generally, state vectors have a length of approximately five [2] [41] [21]. This is a great reduction in dimension but is not so extreme when you consider that, from the evidence in [45], just 3 principal components explains over 95% of the variance in the bond yields. Typically, linear combinations (referred to as portfolios [41]) of bond yields are used as the state variables [17]. In the original paper [2], the authors recommend using five principal components as portfolios.

Why should five principal components be used if the yield curve variation can be explained by 3 principal components only? Our intentions are not to explain just the yield curve, but to predict excess returns. And for our application, more principal components provide additional predictive power [17] [23]. The key point here is that the ACM model also allows a decomposition of the yield curve into a term premium component, and there is no reason that the same variables that explain the yield curve changes, will also describe how that yield curve should be decomposed [62, p.15]. We therefore must include sufficient variables to explain also the variation in term premium. In [17], they show that a single portfolio (linear combination) of yields explains 30-35% of the variance in 1 year excess returns for bonds, which they incorporate into a set of state variables for an affine model [15]. However, the affine model we will use, with 5 principal components of yields as state variables, has been shown [2] to out perform the specification in [15].

4.3 State variables

We set our time step to be one week. This is because we have insufficient quantities of data to train the model to monthly observations; however, using daily observations encounters issues regarding the one day holding period returns for bonds purchased on a Friday, due to the fact that no trading happens over the weekend. Therefore, accurate prices are not available for these days. In order to calculate the excess returns over a one week holding period, we need an estimate of the zero-coupon yield curve, for each cross-section considered in our data, for a range of consecutive maturities, which we decide to be all maturities from 1 week to 10 years. The upper limit of this time window is of course arbitrary, but it is hoped that because no investor can form sensible views about the state of the economy many years in the future, the yield curve beyond 10 years can be inferred from the yield curve up to 10 years. In order to obtain this yield curve we use the Nelson-Siegel-Svensson method [55] where the yield on a zero-coupon bond with time to maturity τ is given by

$$R(\tau) = \beta_1 + \beta_2 \left[\frac{1 - \exp(-\lambda\tau)}{\lambda\tau} \right] + \beta_3 \left[\frac{1 - \exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau) \right].$$

We solve for the five unknown parameters using simplex optimisation [54]. For the state variables themselves, we use the first five principal components of the fitted yield curve. One thing to note is that the principal component analysis is conducted on all the data in the training set and the testing set before it is divided. This is important to ensure that the same linear combinations of yields are used both in sample and out of sample.

4.4 Goodness of fit

The output of our model is of course a time series of yield curves, we plot the input five and ten year yield curve, alongside the output curve as well as the risk neutral output curve in figure 4.1. To evaluate the goodness of fit for our regressions, we consider the average squared residual between the input yield curve and the corresponding output curve. It is natural to expect that in the out of sample period, the quality of fit will slowly degenerate. This is because we cannot reasonably believe that the relationships the model learns will hold true ad infinitum, if indeed they were ever true to begin with. Therefore as time goes on, the true relationship will diverge further and further from the relationship learnt during the training process. This is evident in figure 4.2.

There is clearly a lot of noise affecting the rate of deterioration in the quality of the fit. Therefore, it is not sufficient to evaluate the goodness of fit of the model by considering just the most recent out of sample cross-section. Instead, we consider the average error of all out of sample data points and use that single number to compare how well the model fits to the period of data in consideration. We do this for the most recent 179 business days for South African bonds and show the results in figure 4.3.

We intend to use this averaged out of sample fit error as a measure of how consistent the parameterised yield cross-sections (the input to the model) during the out of sample period are with the entire time series of available data. A low score implies that it is more likely that the market conditions during the out of sample period are similar to the prevalent conditions during the in sample period. This in turn makes it more likely that we can make conclusions based on the cross-sectional yield curve interpolations.

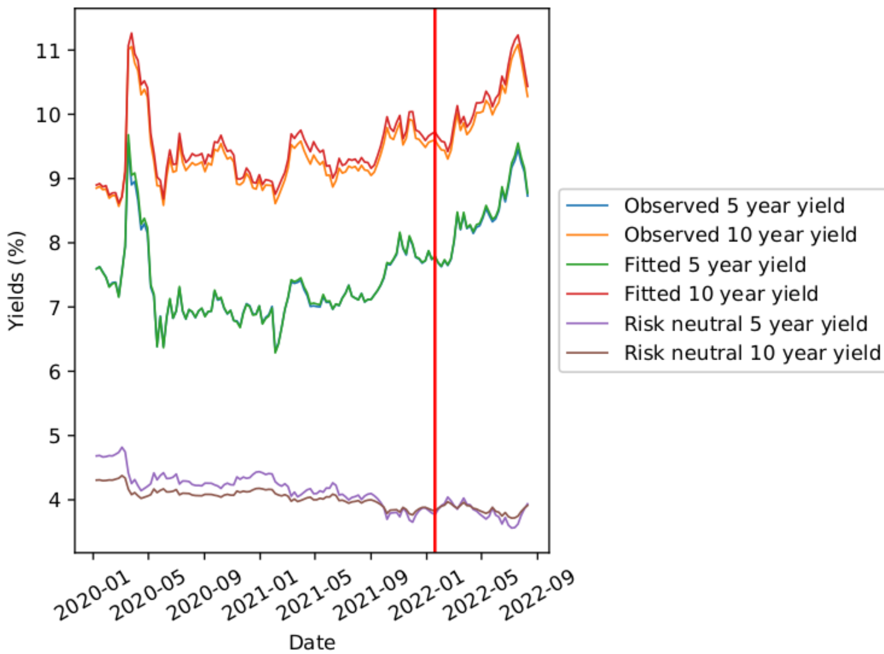
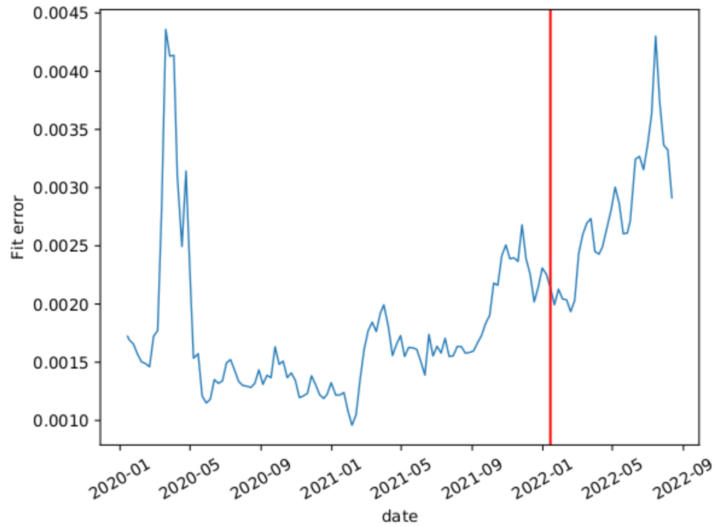


Figure 4.1: The input and output yield curves for the ACM model fit to ZAR bond data. The observed yields are calculated using a fitted Nelson-Siegel-Svensson model. The red line indicates the start of the out of sample period.



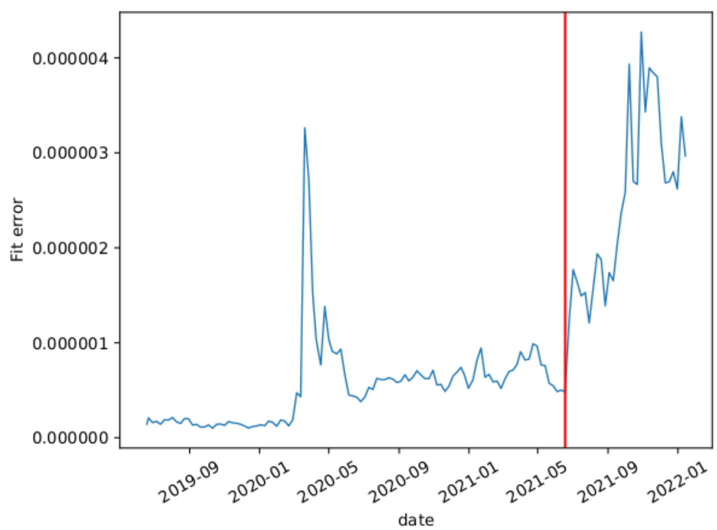


Figure 4.2: The fit quality for both in sample (left of red line) and out of sample (right of red line) for 2 separate fittings of the model to ZAR bond data at different times, compared to the time series of Nelson-Siegel-Svenson snapshot yield curves. The fit quality is the average of the squares of the difference between the input and output yield curves at each instance in time. Unsurprisingly, both plots show a clear spike around the time of the lock-downs caused by the Coronavirus pandemic.

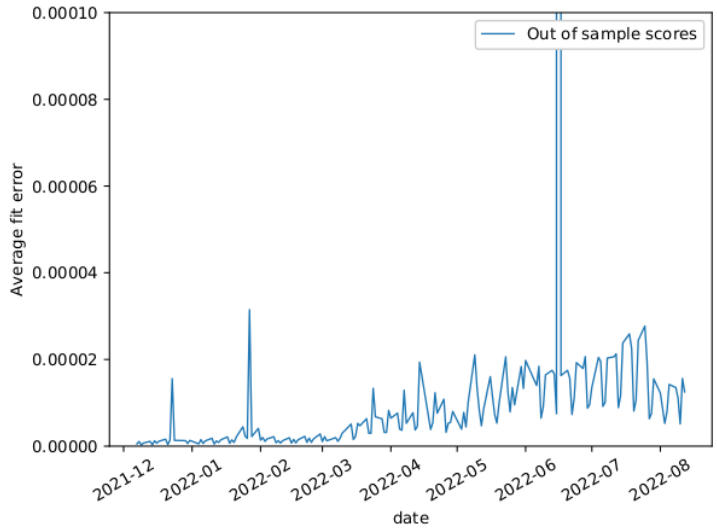


Figure 4.3: The average fit quality for all out of sample cross-sections for ZAR bond data, the time on the x axis is the final date of the out of sample period. Interestingly, the fit quality clearly starts to deteriorate around February to March 2022. This deterioration can likely be attributed to a gradually increasing proportion of the out of sample data being from the period of time following the Russian invasion of Ukraine, whereas the model was fit predominantly during pre-invasion market conditions. The spike during June 2022 highlights a key problem, that the model is numerically unstable. If the training period does not contain enough data, resulting in the wrong relationship being learnt by the various regressions, then the implied bond yields can diverge from the snapshot model's yield curves very quickly.

Chapter 5

Rich-cheap analysis and the ACM model

5.1 Rich-cheap analysis

A bond that is undervalued with respect to the other bonds traded in the market is described as cheap, and if it is overvalued then it is described as rich [26]. Identifying bonds that are trading rich and cheap, or in general, identifying the relative value of all traded bonds, is an important skill for a bond trader. It is important to emphasize that we are only valuing bonds relative to each other, we are not attempting to determine a true value for what the bond is actually worth but merely which bonds are relatively rich or cheap in relation to the other traded bonds. Furthermore, it is not clear weather a bond has a true value in any sense, and the question of what its true value should be is poorly defined. Having identified a cheap bond, a trader may wish to purchase that bond. This purchase can be financed by shorting a nearby bond, or by shorting two bonds, one with a longer maturity and one with a shorter maturity, a trading strategy known as a butterfly [37, chapter 11]. A trader may also use rich cheap analysis to predict the new issue price of a bond, by determining to what extent a bond is likely to trade rich or cheap. This information is then used to adjust the present value of all the bonds future cashflows by an appropriate amount to forecast the bonds initial price. This may be useful in cases where the newly issued bond could become cheapest to deliver for a futures contract, thus effecting the value of said futures contract, as has been known to happen with German Schatz [37, chapter 11].

5.2 Rich-cheap indicators

A common assumption of relative value analysis is that the richness and cheapness of a certain bond is mean reverting [37, chapter 11]; however, for certain bonds this long term mean may be different from 0. We have already discussed the effect of liquidity on bond prices as can be seen particularly in the case of on the run and off the run bonds. In this case, highly illiquid bonds may appear consistently cheap. Nevertheless, a bond can still be considered cheap if it is trading cheaper than it has been in the past, and if this difference is statistically significant then we could suggest that this bond be a good option for a portfolio manager to purchase, if that portfolio manager wished to take a long position in that part of the curve. If there was no mean reversion, then this conclusion would not make sense. This is because we purchase bonds with the intention of selling them later for a profit, and there is no value for us, even if we have bought a highly undervalued bond, unless the market will eventually catch up and bring the price of that bond back into line with its long term mean levels of richness and cheapness. As a rich cheap indicator for a bond x, we quote the market price as a multiple of the Nelson-Siegel-Svensson model implied price i.e.,

$$C^x = \frac{\text{market price}^x}{\text{model implied price}^x}.$$

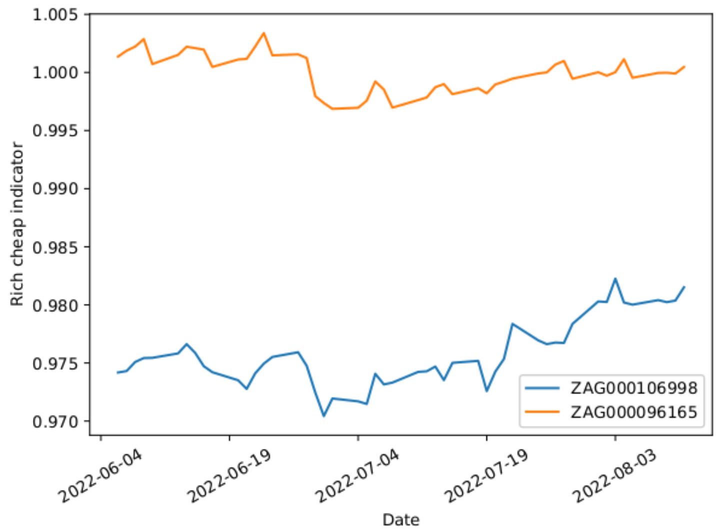


Figure 5.1: The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for two South African bonds over time.

A value greater than 1 means that our analysis indicates the bond is overvalued, and a value less than 1 suggests the bond is undervalued. We show the results for two bonds in figure 5.1.

Unfortunately, it is hard to conclude whether or not the rich cheap indicator shows mean reversion or not, because the time series does not go back far enough, for example, the line indicating the rich/cheapness indicator for the bond 'ZAG000106998', the South African bond with maturity 2030-01-31 bearing an 8% coupon, could be mean reverting back to 0, beginning to diverge from a mean of around 0.975, or anywhere in between these two options. The problem here is that assuming the series does mean revert, its half life of mean reversion [65, p.56] is very long compared to the total length of time that we have available data for. This highlights some of the problems experienced when trying to fit purely statistical models, which we discussed in section 2.2.1. Nevertheless, figure 5.1 does provide at least weak evidence that certain bonds can have rich cheap indicators with long term means different from 1.

We calculate what percentile the most recent rich cheap indicator is within the time series going back 30 weeks of rich cheap indicators for a variety of different bonds being considered. That is, given the time series of rich cheap indicators for a particular bond x up to time t denoted $(C_k)_{k\in\{t-30,t-29,\dots,t-1\}}$ we quote the adjusted richness indicator defined by

$$\tilde{C}^x := \frac{\#\{k \text{ such that } C_k^x < C_t^x\}}{30}.$$

A value greater than 0.5 indicates that the bond is rich, whereas a score less than 0.5 indicates the bond is cheap. We state the results for two different dates in table 5.1 and table 5.2. With the exception of the first bond in table 5.1, the results indicate that all the bonds are rich and therefore should be sold. This result is not particularly believable, and more likely there is some other effect causing the bonds to all increase in value together relative to the prices implied by the Nelson-Siegel-Svensson bond curve. For instance, if liquidity was returning to the market, you would see all prices increasing together as investors demand decreasing compensation for risks due to illiquidity. The problem we have in general when analysing these results is there isn't a long enough time series to see patterns emerging. In this case, we appear to be at a mutual high point in the time series for the adjusted richness indicators, but we have no way of knowing whether this point is globally high, or just part of natural fluctuations of low statistical relevance. On the other

Maturity date	Coupon (%)	Bond ISIN	Adjusted richness indicator
Feb 2023	7.750	ZAG000096165	0.567
Jan 2030	8.000	ZAG000106998	0.967
Feb 2031	7.000	ZAG000077470	1.000
Mar 2032	8.250	ZAG000107004	1.000
Feb 2035	8.875	ZAG000125972	0.867
Mar 2036	6.250	ZAG000030404	0.867
Jan 2037	8.500	ZAG000107012	0.900
Jan 2040	9.000	ZAG000125980	0.867
Feb 2041	6.500	ZAG000077488	0.800
Jan 2044	8.750	ZAG000106972	0.833
Feb 2049	8.750	ZAG000096173	0.867

Table 5.1: A table showing the results of rich cheap analysis conducted on 2022-08-12.

Maturity date	Coupon (%)	Bond ISIN	Adjusted richness indicator
Feb 2023	7.750	ZAG000096165	0.233
Jan 2030	8.000	ZAG000106998	0.500
Feb 2031	7.000	ZAG000077470	0.200
Mar 2032	8.250	ZAG000107004	0.400
Feb 2035	8.875	ZAG000125972	0.600
Mar 2036	6.250	ZAG000030404	0.667
Jan 2037	8.500	ZAG000107012	0.667
Jan 2040	9.000	ZAG000125980	0.700
Feb 2041	6.500	ZAG000077488	0.733
Jan 2044	8.750	ZAG000106972	0.733
Feb 2049	8.750	ZAG000096173	0.667

Table 5.2: A table showing the results of rich cheap analysis conducted on 2022-07-18.

hand, the values in table 5.2 show that the long end of the curve only is overvalued.

5.3 Smoothing of rich cheap indicators

As discussed in chapter 4, we wish to use our results from the ACM term structure model, to validate the rich cheap analysis approach. In particular, we want to detect whether the Nelson-Siegel yield curves, over the period of time in question, show any self consistency. This is an important requirement when you consider that we are using all these yield curves simultaneously to produce the time series from which we take percentiles. To validate a rich cheap indicator we fit the ACM model to the period of time, such that the date we are trying to validate, is the final date of the 30 week out of sample period. As a validation we use that average mean squared error of the yield curves for all of the out of sample period. In figure 5.2, we plot the time series of rich cheap indicators, normalised around their means, for a selection of the bonds in our data set. In figure 5.3, we plot the same series, after the data points have been validated, to validate we take only the points with the lowest 50% fitting error.

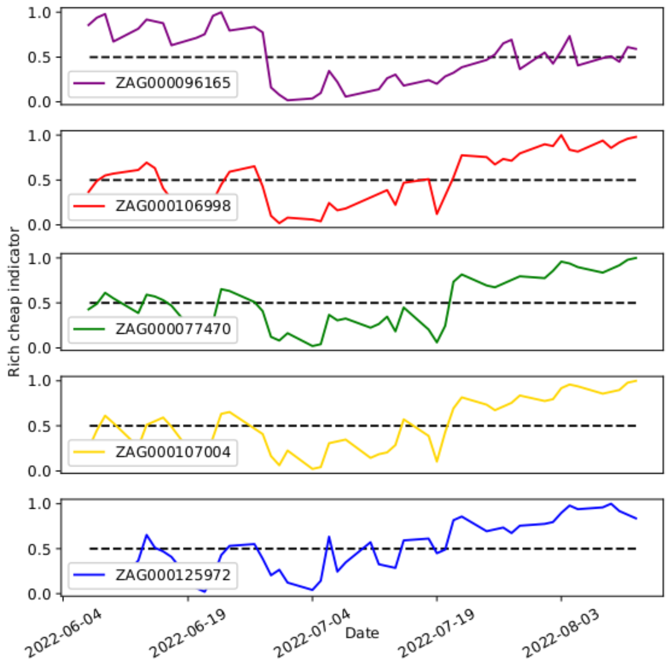


Figure 5.2: The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for a selection of bonds. The salient feature to observe is that the bottom four bonds in the figure, appear to have roughly constant rich cheap indicators up to 2022-07-18 whereafter they all simultaneously become more rich. This conclusion is not easy to swallow; therefore, we hope when we validate the rich cheap indicators, this latter portion will not pass the validation procedure.

It is reassuring to see that the unrealistic simultaneous richness evident in figure 5.2 did not pass the validation procedure, and as such, is not present in figure 5.3.

To compare to our previous results, we alter the definition of adjusted richness indicator by defining a validated adjusted richness indicator as

$$\hat{C}^x := \frac{\#\left\{k \text{ such that } C_k^x < C_t^x \text{ and } C_k^x \text{ passed the validation procedure}\right\}}{\#\left\{k \text{ such that } C_k^x \text{ passed the validation procedure}\right\}}.$$

The most recent data point that passed validation was 2022/07/18 so we give the results of the analysis on that day in table 5.3. The results now imply that the long end of the curve is greatly overvalued, whereas the short end is more fairly priced. This would be interpreted as a strong signal to short the long maturity bonds, which was also present in the unvalidated results from table 5.2.

5.4 Difficulties and criticism

In this section we will discuss some of the issues experienced trying to apply this model, and cast a critical eye over its results and conclusions.

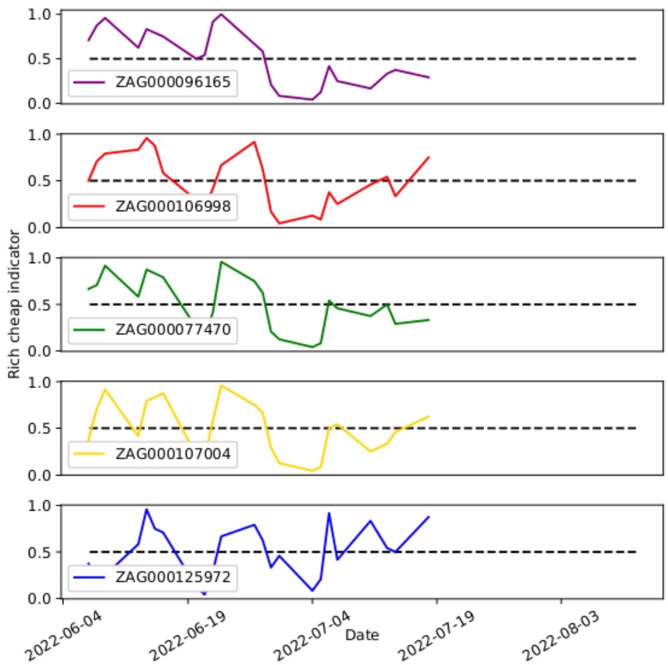


Figure 5.3: The rich cheap indicator, given by (observed market price)/(Nelson-Siegel-Svensson implied price), for a selection of bonds, after the time series has been validated by taking only the observations for days with fitting errors in the lowest 50%. We notice that the period with the unrealistic simultaneous uptick in rich cheap indicator, did not pass the validation procedure. As a result, the time series now runs only to 2022-07-18 in contrast to figure 5.2 that had data all the way to 2022-08-12.

Maturity date	Yield (%)	Bond ISIN	Adjusted richness indicator
Feb 2023	7.750	ZAG000096165	0.292
Jan 2030	8.000	ZAG000106998	0.750
Feb 2031	7.000	ZAG000077470	0.333
Mar 2032	8.250	ZAG000107004	0.625
Feb 2035	8.875	ZAG000125972	0.875
Mar 2036	6.250	ZAG000030404	0.958
Jan 2037	8.500	ZAG000107012	0.958
Jan 2040	9.000	ZAG000125980	0.958
Feb 2041	6.500	ZAG000077488	0.958
Jan 2044	8.750	ZAG000106972	0.958
Feb 2049	8.750	ZAG000096173	0.958

Table 5.3: A table showing the result of rich cheap analysis, with validation procedure, conducted on 2022-07-18.

5.4.1 Emerging markets idiosyncrasies

It is worth bearing in mind that we are trying to conduct our analysis on emerging markets data. These markets are more volatile and less liquid than developed markets, partly because they are exposed to foreign exchange effects, due to the Dollar being the reserve currency for markets globally, and also credit default effects, which are considered negligible in the case of the US. As a result, a more developed market with higher liquidity and trading volume, such as the US bond market, would provide much more data, that would be more stable and reliable. In our case, we are notably lacking in data for bonds with maturities between one and eight years, and it is precisely this period that will carry the most information, as very short term bonds will be very noisy and over longer maturities, investors cannot draw any meaningful conclusions. Given that some of the issues we faced relate to numerical instability, we may have had more success working with a developed economy.

5.4.2 Numerical instability

One of the biggest issue experienced was the numerical instability of the fitting procedure, and its sensitivity to the quality of the input data. For example, the models used in this paper were fit using two years of in sample weekly training data; however, it would be ideal if we could fit to a smaller window (three months say) as discussed below in section 5.4.3. However this approach incurs sever practical difficulties. In figure 5.4 we once again plot the out of sample average fit error for multiple fittings of the model just as in figure 4.3 but with our training window shortened to just 18 months. With only a reduction of 25% in the size of the training set, the errors have exploded. This highlights the practical problems when using a recursive algorithm to generate the yield curve. Namely, if the wrong relationships are learnt and the parameters of the recursion are incorrect, the errors compound on each other very quickly. Principally, the issue is that there are no warning signs that you need to extend the training set or clean the data somehow, the error explosion can show up seemingly from nowhere, for instance the peak value of the spike in figure 4.3 is actually 4.64×10^{23} whereas the rest of the time series has produced an average fit error much less than 1.

5.4.3 Avoidance of regime shifts

The issue regarding the two year training period is more fundamental than just the potential for numerical instability. We use the fit error, between the generated out of sample yield curves and the parsimonious yield curve snapshot models, as a way of validating the snapshot models. By validate, we mean determine whether the snapshot model is consistent with the yield curves fit during the in sample period. However, implicit in this methodology is the assumption that there were prevalent market conditions during the out of sample period. In other words, we are assuming that there has been at most one regime change during the entire period of data used. We are assuming that a poor fit error implies a regime change and a good fit error implies no regime change. The problem with this assumption is that we cannot honestly believe that market conditions remained constant over the 2 year training period. In particular, the size of our training window means it will necessarily include data from the start of the coronavirus pandemic and it would be very surprising if our state variables evolved the same way in the time leading up to the pandemic, as they did during the pandemic. To see how dramatically things could have changed just observe the spikes in the training fit errors in both plots in figure 4.2 around February 2020. This spike will be present if we analyse the in sample fit error of any training set used to produce these outputs. One potential solution to this problem would be to assign a greater significance to more recent observations in the loss function for our regressions during the fitting procedure. In particular, the regressions currently use the sum of squares error function given by [5]

$$E(w) = \sum_{n=1}^{N} (y_n - w'X_n)^2$$

where the coefficients of w are the weights you are training for, X_n are the observations of the input variables and y_n are the corresponding target variables. What we propose as a potential

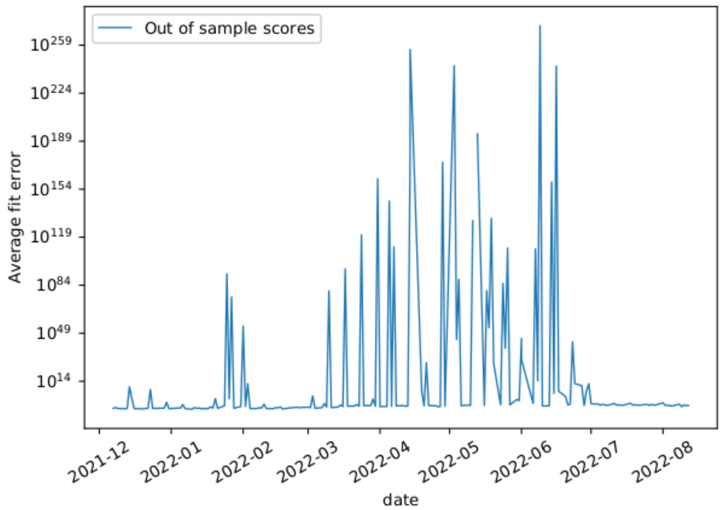


Figure 5.4: The average fit quality for all out of sample cross-sections for ZAR bond data, when the training window was reduced to 18 months only. The time on the x axis is the final date of the out of sample period. Notice that the fit error axis is now logarithmic. When the fit quality is this bad, the model is clearly unusable for any practical purpose.

solution is to introduce a weight term so that our error function then becomes

$$\tilde{E}(w) = \sum_{n=1}^{N} p_n \left(y_n - w' X_n\right)^2$$

for some weights p_n that sum to 1. For example, we could use

$$p_n = \frac{\lambda^{N-n+1}}{\sum_{n=1}^{N} \lambda^n}$$

for some value of λ to be determined empirically. This would result in the weights decaying exponentially and was used in [46] for estimation of covariance matrices [50, p.221]. Exploring this extension is proposed as a future avenue for research.

5.4.4 Validation methodology

Finally, we will discuss the validation methodology in a critical light. The fundamental problem here is that we have had to make compromises due to the fact that we have insufficient, high quality data for bond prices. In order to validate the points in the time series of rich cheap indicators, we removed the bottom 50% of data points by fit error. In the first instance, the choice of 50% is an arbitrary one; however, different values for this cut off limit would all be subject to the same problem, namely, it is not possible for the entire time series we analyse to pass validation. Neither is it possible for the entire time series to be rejected. A better approach would be to decide upon a validation threshold for the maximum fit error we are willing to tolerate, and we reject all points with errors above this threshold. This procedure would be more robust in the event that the entire data series should be either validated or invalidated. Any choice of threshold would have to be tested on a large data set to see how it performs. Looking again at figure 4.3, we can see that the fit error can be divided into 2 regions, a region prior to 2022 when the errors were very small, and a region post 2022 where the errors are much larger. It may therefore make sense for our threshold to be local in a certain sense. Perhaps we could set the threshold to be the maximum out of a weighted average of errors, with higher weights given to more local data points, and some global maximum that we will not exceed. In any case, given the size of the data set we have available for bond data, all these methods will be equivalent to an arbitrary choice of percentage.

Chapter 6

Proposed extensions of the ACM model

In this chapter, we will discuss some of the potential extensions and improvements that we hope to attempt as part of future research.

6.1 Non-spanning variables

Non-spanning variables refers to the possibility that the variables in the yield curve do not span the space of all the risk factors affecting excess returns [62]. In this section, we will consider the use of an expanded set of state variables, including macroeconomic variables, to provide superior predictions versus what was achieved using the yield curve info alone.

The spanning hypothesis is the name given to the principle that the yield curve contains all relevant information about the drivers of the yield curve [3]. If the spanning hypothesis were true, any extension of our state variables with macro variables not derived from the yield curve, would not provide any superior predictive power. However, there has been evidence to suggest that bond returns can be predicted by variables such as inflation [7] [47]; aggregate consumption and real economic activity [11] [7] [40]; and payroll information [60].

Even if you do not assume the spanning hypothesis, the case for macro variables having predictive power is still disputed. For example, Cochrane argues that the enhanced predictive performance of models augmented with macroeconomic factors can be achieved by using only the information contained in the yield curve. He argues in [16] that inflation acts as a proxy for detrending of the yield principal components, that could equally well be achieved by the inclusion of a linear trend factor derived only from yield curve information.

6.1.1 Theoretical challenges to ACM model extension

Another issue with non-spanning factors, in an affine model setting, is the invertibility of the relationship in (3.1.2) for the zero-coupon bond price, which we rewrite here as

$$\begin{aligned} & P_t^{(\tau)} = \exp\left[A(\tau) + B(\tau)' X_t\right] \\ \implies \quad & -y_t^{\tau}\tau = A(\tau) + B(\tau)' X_t. \end{aligned}$$

using (1.2.1) in the last line. In particular, if you assume that the number of factors is kept small compared to the dimension of the space spanned by the observations of the yields, and that the operator $B(\tau)$ is invertible, then the state vector can be recovered from the observations of the yields [62, section 29.4]. Therefore, some restrictions would have to put on B in order to use the affine model machinery developed so far. One approach to solving this problem is to say that only the state variables that correspond to bond yields are used in (3.1.2), but the entire vector of state variables influences the dynamics of the yield variables, i.e., the entire vector appears in place of X_t in (3.2.1) [40]. In particular, (3.2.1) becomes

$$(X_{t+1}, Y_{t+1}) = \mu + \phi(X_t, Y_t) + v_{t+1}$$

where (X_t, Y_t) is a vector formed by the concatenation of the yield variables X_t and the additional variables Y_t (macro or otherwise) that we wish to consider. An alternative to this alteration to the model set up, would be to leave the model the same and argue that the additional factors are important for reducing the measurement error. Recall that the measurement error was a term introduced to account for the possibility that the yields do not behave in the way we have postulated, and in that case, the invertibility argument will not hold. The addition of the non-spanning factors may allow our model to get closer to the true relationship in some sense, without ever obtaining it exactly. It is interesting to note that this invertibility problem is not unique to affine models, and in fact the spanning hypothesis is implied by virtually all macro-financial models [3].

6.1.2 Which factors to choose

In this section we will discuss what factors have been considered in the literature as drivers of yields and term premium. Term premium for different countries shows strong cross-correlation [74] which suggests that global factors may be a strong driver of term premium. In [1], they provide evidence that global factors tend to explain the long-term dynamics of yield curves, whereas the country specific factors explain the short-term dynamics. Further, they show that the curvature of global rates has a strong influence on term premia. For macroeconomic factors, data on inflation [73] and economic growth [74], have been identified as the most significant drivers of term premia. In [74] the authors additionally found the Citigroup country inflation surprise index, the Chicago Board Options Exchange Volatility Index (VIX), and the US term premium to be highly significant. Further, following the work of [22] who show that yield curve liquidity displays some segmentation effects, in [74] they analyse a liquidity indicator developed in [36] and show that it is highly significant when predicting the term premium for the five year portion of the yield curve.

6.2 From data to state variables

The process of reducing the available data to produce state variables is not a trivial one. In this section, we will discuss some of the issues faced during this procedure. It is important to remember that if we wish to use our state variables in the ACM model, they must obey the dynamics specified in (3.2.1), and there is no reason, a priori, that this will be the case for either principal components techniques, or non-linear methods.

6.2.1 Interpretability

A particular issue with using an enlarged panel of data to produce the state variables, is that interpretability of the variables is easily lost [47]. We will discuss this point here, as well as mention some of the potential solutions presented in [47]. For more detail, the reader should consult [47] and the references therein.

Even if you are trying to predict a large number of yields, for instance, all 120 yields all the way from a 1 month tenor to a 10 year tenor going up in months, you are going to run in to problems if you allow the vector of state variables to become too large. This is because there is a lot of redundancy in the yields and as we've already discussed, just three principal components can explain over 95% of the variance. Using a large number of state variables will therefore lead to degrees of freedom issues, meaning that the state variables can overfit to the target yields in a number of ways. Also, the more macroeconomic variables you consider, the larger your chances of finding false positives. Cochrane [16] argues that it is easy to find false positives, and the main guard against this is to ensure that the factors are interpretable, and that the conclusions we are drawing makes sense economically.

Principal components

It is clear therefore that dimension reduction will be required, especially if our starting point is the panel of 131 series used in [47]. However, principal components will not be ideal for this reduction as the factors will be difficult to interpret economically. To get around this problem, one could regress the principal components back onto the raw data, and compute an R^2 value for each regression. The larger this R^2 is, the more we can attribute that principal component to that input variable. Using this technique, Ludvigson and Ng identify 8 factors weighting on, for example, interest rate spreads, prices, real economic activity and money supply.

6.2.2 Non-linearity, machine learning and autoencoders

In this final section, we discuss the use of non-linear methods within a factor structure. The rationale for this extension is based on the work of Bianchi et al. [4], who compared the performance at predicting excess bond returns, of linear regression methods against both deep and shallow neural networks of various specifications. What would be interesting is to use these neural network specifications (or modifications in the same spirit) to produce the state variables out of the panel of macroeconomic information. The output of the neural-network would be the input to the affine model. The choice of loss function that we would try to minimise is also a non-trivial question. The most simple solution would be to train an auto-encoder [66, p.68], that learns how to best describe the inputs in a lower dimensional vector, so that the state variables are as descriptive as possible. The issue with this approach is that reproduction of the input data panel is not the only consideration for our state variables. We also need them to obey the dynamics given by (3.2.1) if we wish to use them as inputs to the ACM model. Therefore, a loss function that incorporates this requirement in some fashion would be sought after.

Bianchi et al. [4], argue that the performance of the network does not increase with the number of layers, and that it is the non-linearities between groups of similar variables that contributes to the increased performance versus linear methods, as supposed to non-linear cross-group relationships. Therefore, they suggest a 'group ensemble' structure, where a neural network with 1' hidden layer containing 1 node is used for each group, where the output of each network is ensembled with the output of an additional neural network for the forward rates (with 3 hidden layers, each with 3 nodes). This approach has a significant advantage, particularly when compared against deeper networks with fully connected layers, in that the output variables can be linked easily back to the closely related data series that the output was mapped from. This ensures that the state variables retain at least some degree of interpretability.

Chapter 7

Conclusion

So far, we have discussed the ACM affine term structure model and applied it to the South African bond market. We then attempted to interpret the results, as buy sell indicators for rich cheap analysis, leveraging the Nelson-Siegel-Svensson model to produce snapshots of the yield curve. Finally, we discussed potential extensions of the model, to allow for non-spanning macroeconomic factors.

The affine models we have applied here are capable of providing a very good fit to in sample market data; however, the fit quality deteriorates very quickly when you try to apply the model to out of sample data. As a consequence of this, the results we obtained when we tried to apply our methods to rich cheap analysis, were very noisy, resulting in them being difficult to interpret. Nevertheless, we did obtain a positive result, namely the successful denoising of the rich cheap indicators in figure 5.3. This result not withstanding, it is still unclear whether the technique has merit, and further research will be required to understand to what extent the methodology works in practice. What is evident however, is that there is merit to the approaches to yield curve modelling that incorporate information from the time series of yield curves, and that models, for example the Hull and White model, that only fit to the current yield curve, may be overlooking important forecasting information.

Bibliography

[1] Mirko Abbritti, Salvatore Dell'Erba, Antonio Moreno, and Sergio Sola. Global factors in the term structure of interest rates. *International Journal of Central Banking*, 14(2):301–340, 2018.

[2] Tobias Adrian, Richard K. Crump, and Emanuel Moench. Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138, 2013.

[3] Michael D. Bauer and James D. Hamilton. Robust Bond Risk Premia. *The Review of Financial Studies*, 31(2):399–448, 09 2017.

[4] Daniele Bianchi, Matthias Büchner, and Tamoni Andrea. Bond risk premia with machine learning. WBS Finance Group Research Paper 252, 2020.

[5] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[6] Fischer Black, Emanuel Derman, and William Toy. A one-factor model of interest rates and its application to treasury bond options, financial analysts journal. 1990.

[7] Michael W. Brandt and Kevin Q. Wang. Time-varying risk aversion and unexpected inflation. *Journal of Monetary Economics*, 50(7):1457–1498, 2003.

[8] Richard A. Brealey, Stewart C. Myers, and Franklin Allen. *Principles of corporate finance*. Wiley, 2003.

[9] Damiano Brigo. *Interest rate models : theory and practice ; with smile, inflation and credit*. Springer, 2nd ed. edition, 2006.

[10] John c. Cox, Jonathan E. Ingersoll, and Stephen Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

[11] John Campbell and John Cochrane. Force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251, 1999.

[12] John Y. Campbell and Robert J. Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

[13] Patrick Cheridito, Damir Filipović, and Robert L. Kimmel. Market price of risk specifications for affine models: Theory and evidence. 2003.

[14] Moorad Choudhry, David Moskovic, and Max Wong. *Fixed income markets*. Wiley, 2 edition, 2014.

[15] John Cochrane and Monika Piazzesi. Decomposing the Yield Curve. 2009 Meeting Papers 18, Society for Economic Dynamics, 2009.

[16] John H. Cochrane. Comments on "robust bond risk premia" by Michael Bauer and Jim Hamilton. Working paper, University of Chicago, 2015.

[17] John H. Cochrane and Monika Piazzesi. Bond risk premia. *American Economic Review*, 95(1):138–160, 1991.

[18] S. D'Amico, D. H. Kim, and Min Wei. Tips from tips: The informational content of treasury inflation-protected security prices. Working Paper 2013-10, Federal Reserve Board, 2010.

[19] Jean-Pierre Danthine and John Donaldson. *Intermediate financial theory*. Academic Press, 3rd edition, 2015.

[20] Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

[21] Antonio Diez de los Rios. A new linear estimator for gaussian dynamic term structure models. Working Paper 2013-10, Bank of Canada, 2014.

[22] Joost Driessen, Theo E. Nijman, and Simon Zorka. Much ado about nothing: A study of differential pricing and liquidity of short and long term bonds. In *Safe Working Paper No 238*, Dec 2018.

[23] Gregory R. Duffee. Information in (and not in) the Term Structure. *The Review of Financial Studies*, 24(9):2895–2934, 05 2011.

[24] Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996.

[25] Eugene F. Fama and Robert R. Bliss. The information in long-maturity forward rates. *The American Economic Review*, 77(4):680–692, 1987.

[26] Financial pipeline. Bond yield and portfolio management: Understanding yield, 2014 URL: https://www.financialpipeline.com/bond-yield/.

[27] finRGN. Arbitrage free vs equilibrium models (frm part 2, book 1, market risk), Sep. 2018 URL: https://www.youtube.com/watch?v=MeIISoA1lfI.

[28] Mark Fisher, Douglas Nychka, and David Zervos. Fitting the term structure of interest rates with smoothing splines. Finance and Economics Discussion Series 95-1, Board of Governors of the Federal Reserve System (U.S.), 1995.

[29] Jean-Sébastien Fontaine and René Garcia. Bond liquidity premia. *The Review of Financial Studies*, 25(4):1207–1254, 2012.

[30] Michael Grabchak and Gennady Samorodnitsky. Do financial returns have finite or infinite variance? a paradox and an explanation. *Quantitative Finance*, 10(8):883–893, 2010.

[31] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[32] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A discrete time approximation. *Journal of Financial and Quantitative Analysis*, 25(4):419–440, 1990.

[33] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, 1992.

[34] Thomas S. Y. Ho and Sang-Bin Lee. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance*, 41(5):1011–1029, 1986.

[35] Hoang, T.A. Khoa, and Robert W. Faff. Is the ex-ante equity risk premium always positive? evidence from a statistical learning method. *Accounting & Finance*, 61(1):95–124, 2021.

[36] Grace Xing Hu, Jun Pan, and Jiang Wang. Noise as information for illiquidity. *The Journal of Finance*, 68(6):2341–2382, 2013.

[37] Doug Huggins. *Fixed income relative value analysis a practitioners guide to the theory, tools, and trades*. Wiley, West Sussex, England, 2013.

[38] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[39] International Capital Market Association, 2020 URL: https://www.icmagroup.org/market-practice-and-regulatory-policy/secondary-markets/bond-market-size/.

[40] Scott Joslin, Marcel Priebsch, and Kenneth J. Singleton. Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, 69(3):1197–1233, 2014.

[41] Scott Joslin, Kenneth J. Singleton, and Haoxiang Zhu. A New Perspective on Gaussian Dynamic Term Structure Models. *The Review of Financial Studies*, 24(3):926–970, 01 2011.

[42] Ioannis Karatzas and Steven E Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 2 edition, 1988.

[43] Orçun Kaya and Jan Schildbach. High-frequency trading, 2016 URL: https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000454703/Research_Briefing:_High-frequency_trading.PDF.

[44] Canlin Li, Andrew Meldrum, and Marius Rodriguez. Robustness of long-maturity term premium estimates, 2017 URL: https://www.federalreserve.gov/econres/notes/feds-notes/robustness-of-long-maturity-term-premium-estimates-20170403.html.

[45] R. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61, 1991.

[46] R Litterman and K Winkelmann. Estimating covariance matrices. In *Goldman Sachs Risk Management Series*, 1998.

[47] Sydney C. Ludvigson and Serena Ng. Macro Factors in Bond Risk Premia. *The Review of Financial Studies*, 22(12):5027–5067, 10 2009.

[48] Barbora Malinská. Time-varying pricing of risk in sovereign bond futures returns. *Finance Research Letters*, 47:102531, 2022.

[49] P. Malliavin, H. Airault, L. Kay, and G. Letac. *Integration and Probability*. Graduate Texts in Mathematics. Springer, 1995.

[50] Lionel Martellini, Philippe Priaulet, and Stephane Priaulet. *Fixed-income securities*. Wiley, 2003.

[51] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[52] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, revised edition edition, 2015.

[53] Sanjay K. Nawalkha, Natalia A. Beliaeva, and Gloria M. Soto. *Dynamic Term Structure Modeling: The Fixed Income Valuation Course*. John Wiley & Sons, 2007.

[54] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965.

[55] Charles Nelson and Andrew F. Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–89, 1987.

[56] Deniz Ozenbas, Michael S. Pagano, Robert A. Schwartz, and Bruce W. Weber. *Liquidity, Markets and Trading in Action*. Springer, 2022.

[57] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[58] Steven E. Pav. *The Sharpe Ratio*. CRC Press, 2022.

[59] Huyên Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Springer Berlin Heidelberg, 2009.

[60] Monika Piazzesi and Eric Swanson. Futures prices as risk-adjusted forecasts of monetary policy. *Journal of Monetary Economics*, 55(4):677–691, 2008.

[61] Juliusz F. Radwanski. Understanding bond risk premia uncovered by the term structure. Working paper, Vienna Graduate School of Finance.

[62] Riccardo Rebonato. *Bond Pricing and Yield-Curve Modelling*. Cambridge University Press, 2018.

[63] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 12th ed. edition, 2019.

[64] Qi-Man Shao, Hao Yu, and Jun Yu. Do Stock Returns Follow a Finite Variance Distribution? *Annals of Economics and Finance*, 2(2):467–486, November 2001.

[65] Ruey S. Tsay. *Analysis of financial time series*. Wiley, 3rd ed. edition, 2010.

[66] George A. Tsihrintzis and Lakhimi C. Jain. *Machine Learning Paradigms Advances in Deep Learning-based Technological Applications*. Springer, 2020.

[67] M. Urbański, M. Roy, and S. Munday. *Finer Thermodynamic Formalism – Distance Expanding Maps and Countable State Subshifts of Finite Type, Conformal GDMSs, Lasota-Yorke Maps and Fractal Geometry*. De Gruyter Expositions in Mathematics. De Gruyter, 2022.

[68] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

[69] Dimitri Vayanos and Jean-Luc Vila. A preferred-habitat model of the term structure of interest rates. Working Paper 15487, National Bureau of Economic Research, November 2009.

[70] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[71] Arthur Warga. Bond returns, liquidity, and missing data. *Journal of Financial and Quantitative Analysis*, 27(4):605–617, 1992.

[72] Alan White and John Hull. Pricing interest-rate-derivative securities. *Review of Financial Studies*, 3:573–92, 02 1990.

[73] Jonathan H. Wright. Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *The American Economic Review*, 101(4):1514–1534, 2011.

[74] İbrahim Özbek and İrem Talaslı. Term premium in emerging market sovereign yields: Role of common and country specific factors. *Central Bank Review*, 20(4):169–182, 2020.