"Exploiting arbitrage opportunities imbedded in the Litterman and Scheinkman term structure model"

Nitzan Regev
Department of Analytic Development
Bank Hapoalim
Tel Aviv, Israel
nitzan.regev@mailpoalim.co.il

Haim Reisman
Department of IE & Management
Technion - Israel Institute of Technology
Haifa 32000, Israel
reisman@ie.technion.ac.il

Gady Zohar
Hedgeware Limited
London, UK

July 16, 2007

Abstract

Litterman and Scheinkman (1991) show that a three factor PCA model of the term structure fits the data in a remarkable way and is useful for hedging. However, this model is not arbitrage free under the assumption that markets are frictionless. That is, there are zero cost portfolios with instantaneous risk free returns which may be different from zero. The paper derives a formula to computes those return, and shows that this formula holds quite accurately in real markets. This is despite of the fact it is inconsistent with no arbitrage in frictionless markets.

1 Introduction

Kaufman, Biewrag and Toves (1983), Gultekin and Rogalski (1984) and others show that the simple Macaulay duration performs at least as well as any other duration measures in bond hedging. Others show that Macaulay duration is very useful in determining hedge ratios in bond markets. Litterman and Scheinkman (1991) and Diebold, Li and Li (2006b) studied a multi-factor version of the above. Diebold Li and Li (2006b) compared two types of factors: Litterman and Scheinkman and Diebold and Li (2006). They obtained improved hedging results using the multi factor model over Macaulay duration. They also found that the portfolio used for hedge gives excess return over the hedged bond.

Reisman and Zohar (2004) show that the prediction of the model for the expected return of zero-cost zero-exposure portfolios may be different from zero. In addition, they also show that despite the existence of theoretical arbitrage opportunities, the model prediction is a good estimator for the return of this portfolio. They used spot-rate data processed by the Federal Reserve. Thus the bonds they used in their calculations were synthetically decomposed from the data. This work examines and extend their results using raw unprocessed bond data.

The difficulties of fitting an arbitrage free affine factor model to the data have been observed by Dai and Singleton (2000). The reason for the success of our approach may be that bond markets have frictions and thus no arbitrage restrictions can be destructive. On the other hand the statistical fit of the PCA is remarkable, thus ignoring its consequences maybe harmful.

We derive an analytical formula for the expected next-period return in coupon-bearing bonds, and estimate the parameters of the formula using historical CRSP data on the US bonds market. We use the suggested formula to theoretically determine the instantaneous expected return of any zero-cost portfolio with zero instantaneous variance. For the same portfolios we then measure the return that is attained in practice. By comparing the long-run actual return of the portfolios with the theoretical return, we find that the derived formula determines the next-period expected return with excellent accuracy.

We also show that our arbitrage formula can explain the expected miss in Diebold et.al. model (2006b). In addition by hedging with different bonds and using optimization we can improve the hedge and get an excess profit. Diebold and Li (2006b) demonstrated that the hedging implications of the Litterman and Scheinkman model are robust to the choice of components. We show the same robustness applies to our arbitrage implications as well.

In section 2 we describe the affine multi-factor models of the term structure, and derive a formula for the expected next-period bond return.

In section 3 we describe the data set used in this paper. We then offer our approach to estimating the spot rates based on the multi- factor model. The estimation procedure is similar to the bootstrapping method. Finally, we calculate the modes obtained by Principal Component Analysis, and the modes obtained by Diebold, Li and Li (2006b).

Section 4 reproduces the comparative risk-management results of Diebold, Li and Li (2006b) and shows that the mean returns of the hedged portfolios studied there are accurately predicted by the model. In addition, we apply the model to obtain better expected payoff in hedging the above target portfolios.

In Section 5 we show how zero-cost portfolios with no factor risk should be constructed. We then introduce simple examples using US Treasury bonds to demonstrate the effectiveness of the methodology. We construct these portfolios and compare their returns to what is expected by the theoretical next-period bond return, and find excellent agreement.

In section 6 we discuss advanced applications. We show how to find zero-cost zero-exposure portfolios with maximal expected returns, so as to exploit arbitrage opportunities in the bond market. We do this by constructing two kinds of portfolios. The first is based on all available bonds, and we find that the average predicted and actual returns are 1.35% for a year. In the Second portfolio presented we seek the maximum return in the presence of transaction costs. The results show that up to transaction costs of 7 cents, arbitrage opportunities can be found.

We conclude in section 7.

2 The Affine Model

We first describe the multi factor-models and the ways to estimate them. Based on those models we derive a formula for the expected next- period bond return.

The spot interest rate at time t for τ years is defined as

$$y(t, t+\tau) = -(1/\tau)ln(B(t, t+\tau)) \tag{2.1}$$

where $B(t, t+\tau)$ is the price at time t of a discount bond with face value of 1 maturing at time $t+\tau$. Affine factor models of the term structure specify the spot interest rate for τ years at time t, $y(t, t+\tau)$, as a linear combination of some F deterministic functions of time to maturity with stochastic coefficients:

$$y(t, t+\tau) = \sum_{i=1}^{F} u_i(\tau) z_i(t). \tag{2.2}$$

The functions $u_i(\tau)$ are called *modes*, and the time dependent coefficients $z_i(t)$ are called *factors*.
The price change at time t of a discount bond maturing at time $t+\tau = T$ is given by

$$\frac{dB(t,T)}{B(t,T)} - y(t,\tau)dt = \sum_{i=1}^{F} -(T-t)u_i(T-t)dz_i(t) + \mu(t,T)dt, \tag{2.3}$$

with

$$\mu(t, t+\tau)dt = 0.5\tau^2 d\langle y(t,T)\rangle + (\partial/\partial\tau)[\tau y(t, t+\tau)]. \tag{2.4}$$

$\langle\cdot\rangle$ denotes the quadratic variation. The second term is the forward rate and the first term is the convexity. The estimation of the quadratic variation in the first term is based on the observed variation. In this work we use monthly data. It turns out that the quadratic variation term has little effect on the results for bonds with maturities up to 10 years (the bonds used to construct the portfolios).
We estimate $\partial y(t,\tau)/\partial\tau$ in the second term as the slope of the term structure at τ. Given N spot rates, in order to compute this slope we must interpolate the term structure. For the results presented in this paper we use cubic spline.
Eq. (2.4) is the drift process. It is obtained by taking an Ito differential of eq. (2.2), assuming the z_i are Ito processes.

Define the price at time t of a bond portfolio with cash flows of C_k at time $t+\tau_k$ until time T as

$$P(t,T) = \sum_k C_k B(t, t+\tau_k). \tag{2.5}$$

The price change is given by[1]

$$dP(t,T) - y(t,\tau)P(t,T)dt = \sum_{i=1}^{F} -D_i(\tau)dz_i(t) + \mu_P(t)dt, \tag{2.6}$$

where

$$Di(t) = \sum_k C_k B(t, t+\tau_k)\tau_k u_i(\tau_k) \tag{2.7}$$

is the exposure of the portfolio to factor i and

$$\mu_P(t,T) = \sum_k C_k B(t, t+\tau_k)\mu(t, t+\tau_k). \tag{2.8}$$

Assuming frictionless markets, our model is arbitrage-free if whenever the $D_i(t)$'s are zero for each i and the value of the portfolio $P(t,T)$ is zero, $\mu_P(t)$ is also zero. Assumptions on the u_i and z_i under which the model is arbitrage free in frictionless markets were studied by Duffie and Kan (1996) and in Reisman and Zohar (2006) in the infinite dimensional case $F=\infty$.

Given a historical set of spot rates $\{y(t_n, t_n+\tau_k); n = 1, \cdots, N; k = 1, \cdots, K\}$, Litterman and Scheinkman (1991) applied Principal Component Analysis (PCA) to obtain the modes that are optimal in the sense that the sum of the squared (in-sample) errors,

$$d_e(t_n,\tau_k) = y(t_n+1, t_n+\tau_k) - y(t_n, t_n+\tau_k) - \sum_{i=1}^{F} u_i(\tau_k)(z_i(t_n+1) - z_i(t_n)) \tag{2.9}$$

is smaller than that obtained for every other choice of modes. That part of the

[1] proof in appendix A

variation of the spot rates not explained by the errors is

$$Q = 1 - \sum_{k,n} \frac{d_e(t_n, \tau_k)^2}{dy(t_n, t_n + \tau_k)2}. \tag{2.10}$$

Most surprisingly, for the data sets examined Q was typically around 98% in the three-factor model, 95% in the two-factor model, and 85% in the one-factor model. The modes obtained can be interpreted intuitively. The first mode is close to a constant, which implies that the one-factor model is in fact the familiar duration model and z_1 describes parallel shifts. The second mode is a monotonic function, implying that the second factor is responsible for the slope of the term structure. The third mode is a convex function, implying that the third factor is responsible for convexity of the term structure. Diebold *et al.* (2006b) repeated Litterman and Scheinkman (1991), replacing the modes with smooth functions that look more or less the same, and obtained similar results. They show that portfolios with a zero-model exposure to all factors have a close to zero out-of-sample variance. However, they do not examine the relationship between the expected return of such a portfolio and its model prediction. In general, the model's prediction regarding the return of zero-cost zero-exposure portfolios may be different from zero, and thus inconsistent with no arbitrage in frictionless markets. The first to argue that in spite of this, the model's prediction of expected returns should be taken seriously were Reisman and Zohar (2004), on the grounds that some low- dimensional PCA model may fit the data in a remarkable way; they showed that for the data they used the model is remarkably accurate. In addition, bond markets are not frictionless. Thus, adding no- arbitrage restrictions when modeling such markets might be destructive, as indicated by Dai and Singleton (2000) and others, who show that low-dimensional arbitrage-free affine factor models do not fit the data very well. The affine factor model with modes obtained by the PCA may allow arbitrage when frictions are ignored. However, when frictions are not ignored this arbitrage may not be fully exploited, as some traders may wish to replace their portfolio with one with the same exposure but a lower expected return.

3 Data Analysis

In the following section we will describe the data set used in this paper. We will also present the procedure in which we estimate the term structure using the multi-factor model suggested above. Finally, we compute and present the modes obtain by the PCA, and the modes obtained by Diebold *et al.* (2006).

Data Set

We use CRSP treasury data from January 1985 to December 2001. The observations periods in the data are organized by month. The computation of the term structure and the hedging on later sections is based on notes and bonds. We use only notes and bonds with a positive face value held by the public, and with bid-and-ask prices. For the bond price we use the average of the bid and the ask prices.

Diebold *et al.* (2006b) and Chambers, Carleton and McEnally (1988) used only bonds maturing on the 15 of February, May, August and November. In this way the authors avoid having too many bonds in the sample, and reduce the number of bonds concentrated at the short end of the term structure. In our work we use additional screening processes. First, we create the term structure (as described below) using all bonds to which apply the rule described above. We then select bonds that are priced correctly - that is, bonds whose bid and ask rates fall within a certain range of the initial term structure. We set this range at 1% above and below the initial term structure. This procedure eliminates any arbitrage opportunities in the bond market, making possible a new, more accurate term structure.
Using all the bonds that fit these rules, we recompute the term structure and in later stages perform hedging.

Term Structure

In order to hedge the proposed modes and to compute the drift formula using real data, we need to estimate the term structure first. There are several ways to do this. All of the estimation methods that we will mention here are based on the multi-factor model given in eq. (2.2).

Diebold *et al.* (2006) used the bootstrap method to compute zero bond yields from available bonds. They computed their three-factor loadings (explained later in this section), and run a regression of the zero yields on the factor loadings (modes). A second approach is to estimate the bond price. The estimated price is based on discounting each coupon using eq. (2.2). Therefore one needs to find factors that bring the sum squares of price errors to a minimum. Using estimated prices can present an accuracy problem, as we estimate the price and not the spot rates directly. The fit problem is mainly at the short end of the term structure, because a small fit error in the bond price can be translated to a bigger error in the estimated spot rate [2].

The third approach, which we use here, is to estimate the term structure by directly computing the spot rates. Again, we use eq. (2.2). While Diebold *et al.* (2006) factors loadings can be used as modes, we chose to use different modes. We set eight modes, meaning that in eq. (2.2) $F = 8$. Each mode is a vector of eight components. The u_i is set to be the natural base, such that $u_1(\tau) = (1,0,0,0,0,0,0,0)$, $u_2(\tau) = (0,1,0,0,0,0,0,0)$ and so on. There are eight components in each factor, and we set one maturity for each component. The maturities we use are [3M,6M,1Y,2Y,3Y,5Y,7Y,10Y] (M for month, Y for year). As mentioned, we seek the minimum sum squares of spot rate errors. The spot rate errors are computed for the selected bonds maturities. At each time t we solve the following optimization problem, when our variables are z_i from eq. (2.2)

$$\min_{z_i(t)} \sum_{j=1}^{M} \{ y_j(\widetilde{t, t+\tau}) - y(t, t+\tau) \}^2. \tag{3.1}$$

M is the number of all available bonds at time t. We compute the term $y_j(\widetilde{t, t+\tau})$ from the data in the following way:

$$y_j(\widetilde{t, t+\tau}) = \{ \frac{1}{\tau_K} log[\frac{p_j - \sum_k^K \exp(-y(\tau_k)\tau_k c_j)}{c_j + 100}] \}, \tag{3.2}$$

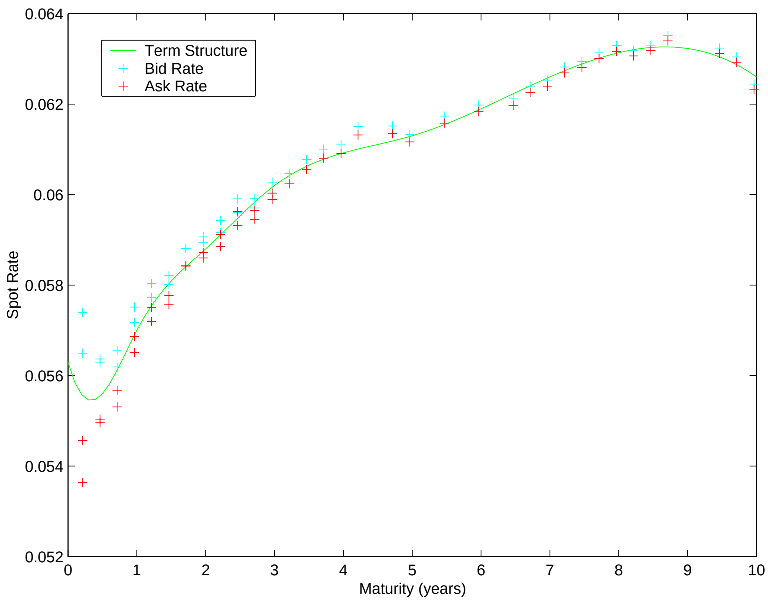
where c_j is the coupon of bond j, K is the number of coupons in bond j, τ_k is the time to maturity of coupon k in bond j, and p_j is the price of bond j.

[2] We estimate the term structure using this procedure as well, improving it by adding a weight for each price error. As modes in eq. (2.2) we use Diebold *et al.*'s (2006) factor loadings. More details and final results are presented in appendix C.

This means that for each bond we compute the spot rate of the last coupon, where earlier coupons are discounted by eq. (2.2). The firsts bonds pay only one coupon, the second bonds pay two coupons and so on. This procedure is very similar to bootstrapping. Using the proposed minimization problem we estimate the spot rates that best fit the spot rate computed in eq. (3.2) at the maturity (time T) of each available bond. It should be noted that in the data we do not necessarily have coupons being paid on the exact maturities we set for u_i and $z_i(t)$. Therefore we use cubic spline (Matlab) interpolation in order to compute the spot rates on different maturities.
By using the last estimation procedure we can accurately estimate the term structure, as shown next.

The following figure show the term structure (green line) estimated on May 1997. We also present the bid (cyan points) and ask (red points) spot rates of all bonds maturing on the 15th of February, May, August and November. The X-axis is the maturity in years, and the Y-axis is the yield.

Figure 3.1: Term Structure computed for May 1997



In figure 3.1 we present the same term structure data presented by Diebold and Li (2006). When comparing their term structure figure and Figure 3.1, we notice that the term structure behaves in a similar way, and the yields are in the same range. Appendix B shows all the term structures dates presented by Diebold and Li (2006). Again, the term structures behave in a similar way, and the yields are in the same range.

PCA Modes

PCA is short for Principal Component Analysis. The PCA is a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component captures most of the variability of the market data, and each succeeding component captures as much of the remaining variability as possible.
We set $u_i(\tau)$ as the orthonormal principal direction determined by applying standard PCA to a sequence of vectors $(dy(t, \tau_1), \ldots, dy(t, \tau_N)) \in \mathbb{R}^N$, and then defining dz_i, $i = 1, \ldots, F$ as the orthonormal projections of $dy(t, \cdot)$ onto $u_i(\cdot)$

$$dz_i(t) = \sum_{j=1}^{N} dy(t, \tau_j) u_i(\tau_j). \tag{3.3}$$

Because the only assumption made is eq. (2.2), the statistical properties of $z_i(t)$, $i \in \{1, \ldots, F\}$ can be almost anything. Using PCA we obtain the affine decomposition that best explains the observed variance of dy among all decompositions with F factors. In addition, PCA has the significant advantage of requiring no ad-hoc assumptions on the dynamics of the process. Litterman and Scheinkman (1991) and Reisman and Zohar (2004) showed that up to 98% of bond variation can be explained by the first three principal components.
It should be noted that, in contrast to other models, our use of the principal component model does not impose a no-arbitrage constraint. Using a relaxed constraint, the fitness of our estimators is at least as good as with other models. Relaxing this constraint makes it possible to find arbitrage opportunities.

We compute the leading principal direction based on the learning data from 31.01.1985 until 30.01.1992, and the estimated term structure for that period. We apply the PCA on the vectors $(dy(t, \tau_1), \ldots, dy(t, \tau_N))$ for all t in the learning data, and determine the principal directions $u_i(\cdot)$. The following table shows the decomposition from the data. The first column shows the maturities we set, and the remaining columns represent the principal components. The components columns describe the move in dy as the result of a one-unit change in a certain factor z_i.

Table 3.1: Factor Loadings
We present the value for each factor loading (U_i) at each maturity.

Maturity:	U_1	U_2	U_3
3M:	-0.262	-0.6064	0.6362
6M:	-0.3047	-0.4435	-0.1359
1Y:	-0.3587	-0.2655	-0.3492
2Y:	-0.3802	-0.0456	-0.3146
3Y:	-0.3950	0.0774	-0.2493
5Y:	-0.3812	0.2644	-0.0601
7Y:	-0.3698	0.3796	0.1538
10Y:	-0.3566	0.3783	0.5162

In PCA decomposition, factors are listed in the order of their importance, based on how much of the total variation they explain. In the next table we show the standard deviation and the percentage of variance explained by the factors. The last row in the table is the cumulative variance explained.

Table 3.2: Factors Std, explained percentage of total variance and cumulative percentage of the explained variance We report two statistical properties of the factors, and the overall variance explained by the modek with i factors.

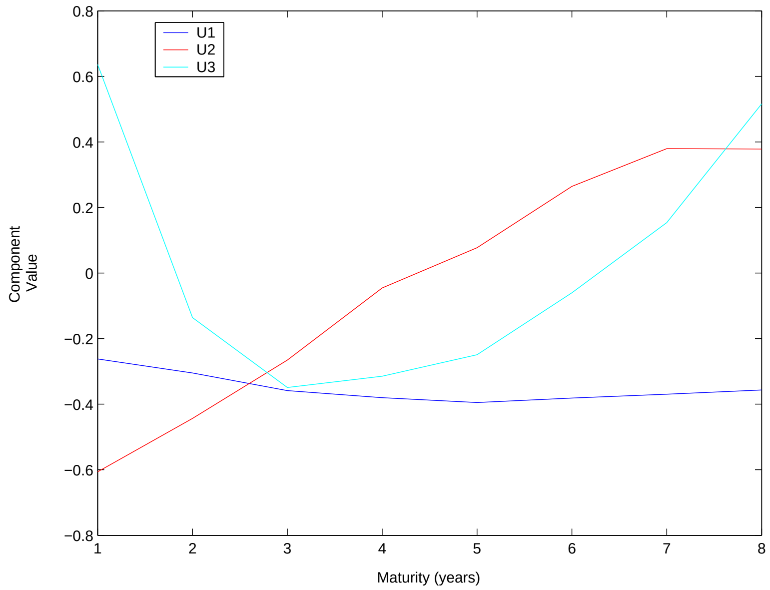
Measurement :	U_1	U_2	U_3
Std:	0.0852	0.0251	0.0138
Var Explained:	88.8%	7.71%	2.32%
Cumulative Var :	88.8%	96.51%	98.83%

We learn from Table 3.2 that the first factor explains about 88.8% of the total variance in the term structure. The second factor accounts for an additional 7.71% of the variance, and the third factor contributes an additional 2.32%. Therefore more than 98.8% of the total variance in the term structure can be explained by as few as three factors.

To help explain the role of each of these three factors, the factors are plotted in the next figure. The X-axis is maturity in years and the Y-axis is the value of the components.

Figure 3.2: The first three principal components $U_1(\tau), U_2(\tau), U_3(\tau)$

We present the value of each principal component as a function of its maturity.



In figure 3.4, the first factor (blue line) is seen to be almost constant as a function of maturity. Therefore the first factor is responsible for parallel shifts of the term structure, and so plays the same roll as duration. The second factor (red line) corresponds mainly to the tilt of the curve. The third factor (cyan line) is responsible for the curvature of the term structure.

These principal directions seem to agree with those found in earlier works, specifically Litterman and Scheinkman (1991) and Reisman and Zohar (2004). It is important to note that the u_i for $i = 1, 2, 3$ depend very little on the time interval over which they are estimated. This is not the case for principal directions of a higher order. Reisman and Zohar (2004c) showed that the stability for those principal directions is very weak (changes rapidly in time). Therefore one should avoid using models with more than three factors.

Summarizing the above, the model should consist of at least two factors, thereby capturing more than 96% of the variation in the term structure. But the model should not consist of more than three factors. If more than three factors are used, the principal directions will be sensitive to the selection of the learning period.

Diebold, Li and Li Modes

Diebold, Li and Li (2006b) used the three-factor model suggested by Nelson and Siegel (1987). They reformulated Nelson and Siegel's (1987) original model to:

$$y(t,\tau) = f_{1t} + f_{2t}\frac{1-\exp(-\lambda\tau)}{\lambda\tau} + f_{3t}(\frac{1-\exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau)). \tag{3.4}$$

The authors offered an economic explanation for the parameters f_{1t}, f_{2t} and f_{3t} (which we call factors). The first parameter is a level factor because its loading is 1. The interpretation of the second parameter is slope, because it affects primarily short term yields. The loading is $\frac{1-\exp(-\lambda\tau)}{\lambda\tau}$, which starts at one and decays monotonically to zero. The third factor is curvature one. The factor f_{3t} has the loading $(\frac{1-\exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau))$, which starts at zero, increases and then decays. Estimation of the factors is by Ordinary Least Squares (OLS) using the factor loadings and the yield curve. Before using OLS one should find λ, in order to compute the factor loadings. In Nelson and Siegel's (1987) model the parameter was λ_t and could change over time. The importance of λ is that it determines the maturity at which the curvature reaches its maximum. As described above, the third factor loading is responsible for the curvature. One known fact from historical term structures is that the curvature reaches its maximum between maturities of two to three years. Based on this, Diebold and Li (2006) found out that $\lambda = 0.0609$ gives the maximum value to the third factor at a maturity of 30 months. They also discovered that a small change in λ does not greatly affect the results.
One might argue that the use of 30 months based on historical observations can make the results in-sample. Here, however, when we use Diebold *et al.* (2006) factor loadings, we consider the results as out-of-sample.
In this paper the maturities are set as years, therefore $\lambda = 0.7308$[3].

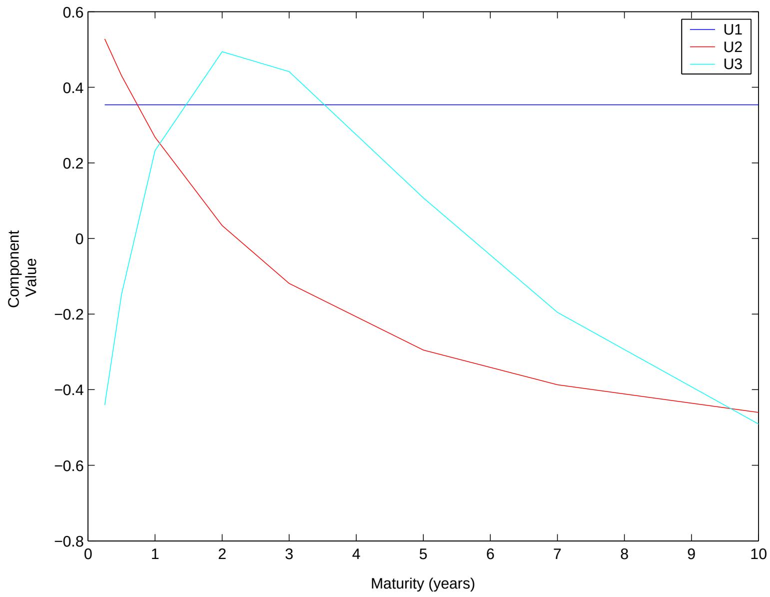
In Figure 3.5 we present Diebold *et al.*'s (2006) factor loadings. The computation of each factor loading in their model depends only on λ and τ. Here we compute the

[3]We also use different values for λ.We present the results for the 6th section in appendix D

authors' modes using Gram-Schmidt, because the derived expected return formula (eq. (2.4)) is based on orthonormal vectors.

Figure 3.3: Diebold *et al.* Factor Loadings

We present the value of each factor loading as a function of its maturity.



From figures 3.4 and 3.5 we see that the three factors are almost the same, with the only difference being a minus sign.

In this section we described the data and the term structure estimation method. We also presented the modes computed both by PCA and Diebold *et al.*'s (2006) factors loadings. From the last two figures we see that the modes in the two methods are the same. This should not come by a surprise to us. First, Diebold *et al.* (2006b) proved that their three modes span approximately the same space as the first three leading principal directions. Second, the first mode in Diebold *et al.* (2006) is level, a result we achieve using the PCA. The construction of the second mode creates a tilt in the curve, which it is likewise the result we derive from the PCA. The third mode in Diebold *et al.* (2006) is created to add curvature to the term structure. This is also the role of the third principal direction.

4 Comparing Risk Management Methods

In the previous sections we presented the multi-factor model. As shown, there are different ways to estimate the modes in this model. Diebold *et al.* (2006b) explored the use of the multi- factor model for risk management. In their paper they focused on three risk management methods. Their purpose was to show that the use of multi factor model for hedging outperforms other chosen hedging mechanisms.

The first method they used for hedging was traditional Macaulay duration. Macaulay duration hedging is appropriate only for the level mode, meaning it is equal to the first factor loading in Diebold *et al.*'s (2006) model. It was shown in earlier papers that duration-based hedging works well only when the yield curve undergoes small parallel shifts.

The second risk management method discussed was Polynomial Duration. This method consists of three elements, each of which is used for hedging. The polynomial duration vector for each bond is $(\sum_{k=1}^{K}\omega_k\tau_k, (\sum_{i=k}^{K}\omega_k(\tau_k)^2)^{\frac{1}{2}}, (\sum_{i=k}^{K}\omega_k(\tau_k)^3)^{\frac{1}{3}})$, where τ_k is time to maturity of coupon k and ω_k is the ratio between coupon k's present value to the bond price.

The final hedging method was the multi-factor model. By using the multi-factor model one can hedge against level risk, slope risk, and curvature risk.

In order to compare the three methods, Diebold *et al.* (2006b) needed to define the bond duration measure for the last method. To do so they use eq. (2.2), but replacing $z_i(t)$ by the term f_{it}. This means, again, that the zero-yield curve is linear in some arbitrary factors f_{it}. The duration measure vector based on their three-factor model of the yield curve captures the form

$$(Du_1, Du_2, Du_3) = (\sum_{k=1}^{K}\omega_k\tau_k, \sum_{k=1}^{K}\omega_k\frac{1-\exp(-\lambda\tau_k)}{\lambda}, \sum_{k=1}^{K}\omega_k(\frac{1-\exp(-\lambda\tau_k)}{\lambda} - \tau_k\exp(-\lambda\tau_k)))$$

for each bond. The symbols are as before.

The introduced methods are used as hedging tools in a practical bond portfolio management context. The purpose of hedging is to protect the value of the bond portfolio against interest changes by controlling its components, such as the duration vector.

In practice, in order to hedge one needs to select a bond as the target asset and to construct a matched bond portfolio. The construction of the portfolio is such that its duration measure is the same as that of the target asset. If the duration measures are the same, then the target asset and the bond portfolio should have the same return. After constructing the portfolios and setting the target asset, the authors computed the realized returns. The difference between the portfolio-realized return and the target asset-realized return is the hedging error. The authors compared three different types of statistical measurement based on the hedging error. The first is Mean Absolute Error (MAE), while the second is the Standard Deviation (Std) of the hedging error. The smaller those measurements, the better the hedge. The authors also computed the mean hedging error (ME). When the mean error is positive the portfolio gains more return than the target asset. This is a desirable property, and we would like to show that it is parallel to Reisman and Zohar's (2004) theorem for excess yields. If indeed this is a parallel property, then we have an explanation for the non-zero mean hedging error, and can exploit it for arbitrage opportunities.

Diebold *et al.* (2006b) solved a similar minimization problem to the one we show next for each hedging method in order to compute the realized returns. The portfolio return depends on the β (shares) of each bond.

$$\min_{\beta_j(t)} \sum_{j=1}^{M} (\beta_j(t))^2$$
$s.t$
$$\sum_{j=1}^{M} \beta_j(t) Du_{ij}(t) = Du_{iTarget}(t)$$
$$\sum_{j=1}^{M} \beta_j(t) P_j(t,T) = P_{Target}(t,T)$$

The symbols are as before except that Du_{ij} is the duration measure i of bond j. In the previous sections we called this $D_{ij}(t)$. The difference between our work and Diebold *et al.* (2004) is that here we use β as the number of shares for each bond, and they used it as a specific bond weight in the portfolio. In this section, when we hedge using the multi-factor model we take Diebold *et al.*'s (2004) factor loadings and not the modes obtained by the PCA.

The main purpose of this paper is to investigate the adjustment of the drift theorem to coupon-bearing bonds. Therefore we will add to the tables presented by Diebold *et al.* (2006b) one more column, which presents the theoretical mean hedging error. We compute the theoretical mean hedging error for each hedging mechanism, using eq. (2.4), meaning that we compute the target asset and the portfolio drifts.

The following tables show the comparison between the different hedging mechanisms. Column one presents the hedging methods, including duration; two and three factors for the multi-factor model; and polynomial duration (PD). Column two is the mean absolute error, column three is the standard deviation, column four is the mean hedging error, and the last column is the predicted mean hedging error.
In Tables 4.1 and 4.2 we hedge five-year bonds against nine different bonds with maturities of one to ten years. The difference between the two tables is in the portfolio holding period. In Table 4.1 we hold the portfolio for one month, meaning that each month we select new bonds that are closest to the maturities we set, and construct a hedged portfolio. In Table 4.2 we hold the portfolio for six months. In order to keep a perfect hedge we rebalance the portfolio according to the duration measure every month during the holding period, meaning we rebalance only the selected bonds. We present monthly results in the following tables.

Table 4.1: Description of the statistics measurement- holding portfolio one month
We present the hedging performance for the three hedging strategies. We measure the performance by the mean absolute error (MAE), standard deviation of the error, and predicted and actual mean error (ME). The hedging is done every month.

Hedging Method:	MAE	Std	ME	Predicted ME
Duration:	0.0013	0.00155	0.000194	0.000196
2 Factors:	0.00115	0.00136	0.000242	0.000202
3 Factors:	0.00097	0.00117	0.000158	0.0009
PD:	0.00096	0.00116	0.000254	0.000172

Table 4.2: Description of the statistics measurement- holding portfolio six months We present the hedging performance for the three hedging strategies. We measure the performance by the mean absolute error (MAE), standard deviation of the error, and predicted and actual mean error (ME). The hedging is done every six months.

Hedging Method:	MAE	Std	ME	Predicted ME
Duration:	0.0013	0.0016	0.00024	0.000193
2 Factors:	0.00011	0.0013	0.000284	0.000204
3 Factors:	0.000856	0.0010	0.00017	0.00011
PD:	0.000913	0.0011	0.000252	0.000181

From Tables 4.1 and 4.2 we see that the two-factor and three- factor hedging methods outperform regular duration hedging. This conclusion is based on the fact that these methods have smaller mean absolute hedging errors and smaller standard deviations. The mean hedging error is positive for all the hedging mechanisms. This is true for both the holding periods. When comparing the multi-factor model and the polynomial duration we find similar results, meaning it is not obvious which method is better. The results in the tables are similar to those found by Diebold *et al.* (2006b). We will also investigate the theoretical mean error. The predicted mean error is smaller than the actual one. Assuming that the predicted error has normal distribution, we can create hypothesis tests. We test whether the predicted mean hedging error is equal to the error in practice (H_0), or whether the two mean hedging errors are different (H_1). Up to a confidence level of 95% we cannot reject the null hypothesis, in both the tables and for all the hedging mechanisms. Therefore we can conclude that the predicted mean hedging error is not different from the actual error. The ratio between the predicted mean error and the actual one ranges from 0.60 to 0.99. One can conclude that more than 60% of the hedging error can be predicted. As we add more factors, the accuracy of the prediction falls, possible because more shares are used from each bond. This is true for both of the holding periods.
In Table 4.3 we present a five-year bond hedge against 21 different bonds for a holding period of one month.

Table 4.3: Statistics Measurement using 22 bonds
We report the hedging performance for the three hedging strategies. We measure the performance by the mean absolute error (MAE), standard deviation of the error, and predicted and actual mean error (ME). The hedging is done every month, and we use more (22) bonds for the hedging.

Hedging Method:	MAE	Std	ME	Predicted ME
Duration:	0.00148	0.0018	0.0002	0.000162
2 Factors:	0.00126	0.00148	0.000275	0.000222
3 Factors:	0.00093	0.00112	0.00022	0.00015
PD:	0.00097	0.00116	0.000274	0.00019

The conclusions from Table 4.3 are the same as from Table 4.1, although the mean absolute error and the standard deviations are smaller in Table 4.1. Using the same hypothesis test as before, we find that up to a confidence level of 95% we cannot reject the null hypothesis. The ratio between the predicted and the actual error starts a little higher, from 0.69-0.8. It is also noteworthy that the predicted mean error in Table 4.3 is higher for the three-factor model and the polynomial duration than in Table 4.1. For the two-factor model it is almost the same.
In Table 4.4 we hedge one bond with maturity of five years against two bonds, one with maturity of two years and one with maturity of eight years. This scheme helps us to check whether we can use the drift formula for simple duration hedging. By simple we mean that we use only one factor and only two bonds from both sides of the five-year maturity bond.

Table 4.4: Duration hedging statistics measurement

newline We present the hedging performance for duration hedging. We measure the performance by the mean absolute error (MAE), standard deviation of the error, and predicted and actual mean error (ME). The hedging is done every month, and using three bonds.

Duration Holding period:	MAE	Std	ME	Predicted ME
1M:	0.0018	0.0021	0.00015	0.000207
6M:	0.00019	0.0023	0.000283	0.000218

One can learn from this table that the hedge is worse, with both the MAE and the Std higher than in the other tables (using more bonds). But we can still predict the mean hedging error fairly accurately. We base this conclusion on the fact that we cannot reject the null hypothesis.

Finally, our intention in this section was not to duplicate Diebold *et al.* (2006b). We used the guidelines of their paper and found similar results, suggesting that the presented multi-factor model is a good risk-management instrument. The difference in our work is that we show why this hedging mechanism gives better results, and offer an explanation for the mean hedging error based on the drift theorem. We also checked a simple duration hedging and found that we can predict the mean hedging error with fair accuracy. It should be emphasized, again, that mean hedging error is a positive property that can be used to create arbitrage opportunities.

5 Simple Examples: Portfolio Construction

In the following section we will show how to construct a portfolio with no exposure to risk factors and at zero cost. Using this portfolio we will demonstrate simple examples for the predictability of the drift formula.

The value of a portfolio at time t is defined as $V \equiv \sum_j \beta_j(t) p_j(t,T)$, when $\beta_j(t)$ is the weights of $bond_j$ in the portfolio. The instantaneous change in V is:

$$dV = \sum_{j=1}^{M} \beta_j(t) \frac{dp_j(t,T)}{p_j(t,T)}. \tag{5.1}$$

In order to neutralize the instantaneous risk-free rate, $r(t)$, the value of the portfolio at time t is equal to zero. Therefore it satisfies the requirement that $V(t) = \sum_{j=1}^{M} \beta_j(t) p_j(t,T) = 0$. Using eq. (2.6) with eq. (2.2) we can immunize the exposure to risk by eliminating $z_i(t)$, due to the fact that $z_i(t)$ depends only on t and not on the maturities as u_i does. Hence, the value of any portfolio is independent of z_i if and only if

$$\delta_i \equiv \sum_{j=1}^{M} \beta_j(t) \frac{D_{ij}(t)}{p_j(t,T)} = 0. \tag{5.2}$$

Any portfolio that satisfies eq. (5.2) for all $i \in \{1, \ldots, F\}$ has no exposure to risk factors according to the multi-factor model. The u_i can be principal directions from the PCA or Diebold *et al.*'s (2006) factor loadings. We need $F+1$ bonds to construct a risk-free portfolio with value of zero. If we wish to hedge the portfolio against a target asset we need one more bond. Given at least $F+2$ bonds, such a portfolio is simply a solution of a set of linear equations.

With the assumption that there is no residual risk, eq. (2.6), eq. (2.2) and eq. (5.1) imply that the excess return of a zero-cost portfolio with no exposure to risk factors is:

$$\mu(\beta) := \sum_{j=1}^{M} \beta_j(t) \mu_j(t, t+\tau). \tag{5.3}$$

This equation gives the definition for the theoretical next-period portfolio return at time t of a zero-cost portfolio with no exposure to risk factors. Under our constraints the portfolio is instantaneously riskless. However, in practice there may be some residual noise. Therefore the terminology "risk-free" refers to a portfolio that entails no exposure to risk factors, but may show some residual risk in reality, a matter that we look at next. To understand this we need to compare the theoretical and actual variance. The theoretical variance of the change in the portfolio value is by definition

$$Variance\{\sum_{j=1}^{M}\beta_j(t)\frac{dp_j(t,T)}{p_j(t,T)}\}. \tag{5.4}$$

Simplifying this equation we see that eq. (5.4) is equal to

$$\sum_{l=1}^{L}\frac{\sum_{i=1}^{F}\gamma_{il}^2\sigma_{zi}^2}{L} \tag{5.5}$$

where

$$\gamma_i = \sum_{j=1}^{M}\beta_j(t)\frac{D_{ij}(t)}{p_j(t,T)}. \tag{5.6}$$

We compute σ_{zi}^2 from the observed data, and L is the number of months (periods) being analyzed. F is the number of all factors, meaning that in the PCA not only the first three factors are used. It should turn out that for the factors used for hedging the $\gamma_i = 0$ $(i \in \{F\})$.

In all the following examples the portfolios do not include cash, meaning they are zero-cost portfolios. Therefore the results are completely insensitive to changes in the instantaneous rate $r(t)$. Using the estimated leading principal directions $u_i(\tau)$ or Diebold *et. al.*'s (2006) factor loadings, we compute for every t the drift $\mu_j(t,\tau)$ for the selected bonds. Finally, a zero-cost portfolio is constructed with no exposure to factor risk, and the existence of nonzero excess returns is shown. These returns, which may vary substantially among such portfolios, are in excellent agreement with those predicted by the theory. We can construct these portfolios since the number of leading factors is smaller than the number of traded securities. Therefore there is a continuum of zero-cost instantaneous portfolios with no factor risk.

At first we will duplicate Reisman and Zohar (2004) and Diebold *et al.*'s (2006b) procedure, meaning we will select bonds with maturities closest to 6M, 1Y, 2Y, 3Y, 5Y, 7Y, and 10Y (Diebold *et al.* (2006b) chose different maturities). At each trading point, which is the time we set to construct a new portfolio, we sell all holdings in the chosen maturities and enter a new position in the bonds with maturities closest to the above. Note that the same bond may be used at different trading points. The shortest trading point is one month as the data are organized, but we also use six months as a trading point. If the trading point is longer than one month the portfolio is rebalanced every month (with the same bonds) to maintain the constraints. Clearly, in practice one should devise schemes that are less expensive in terms of rebalancing costs.

In order to determine the actual return at each trading point of a certain portfolio we need to find the actual return of each bond and multiply its weights, $\frac{p_j(t+1,T)-p_j(t,T)}{p_j(t,T)}\beta_j$, Summing all the bonds in the portfolio gives the portfolio return.

When analyzing empirical results the magnitude of excess gains is quantified by summing up monthly gains without discounting. In other words, the gains do not reinvest, and so the plots do not represent growth of wealth (which would be exponential). Instead, one should regard the slope of these plots as the instantaneous rate of profit from the portfolio while restarting every trading point from a zero-net position.

The purpose of the following examples is not to achieve big excess returns (see advanced application), but rather to show that there are excess returns in the bond market and that they are substantially nonzero. We will also demonstrate the good fit between the drift in theory and the excess returns gain in practice.

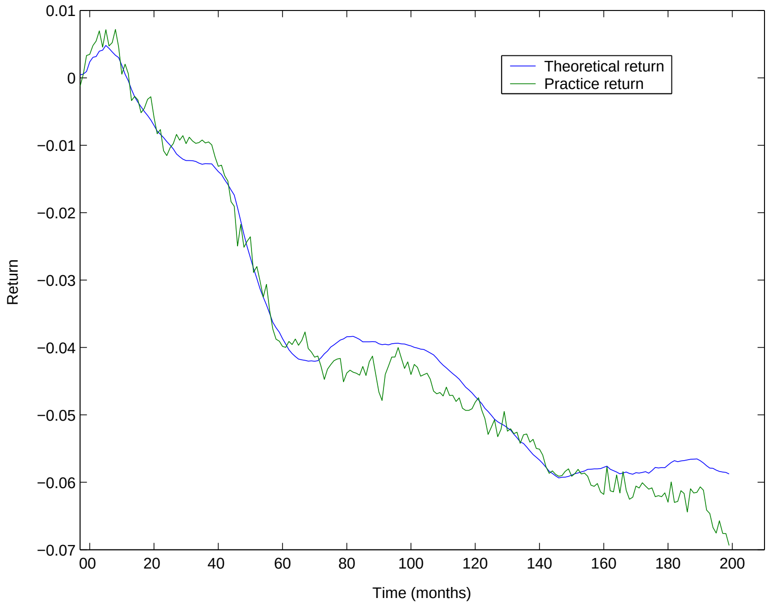
In this section we use the leading principal directions $u_i(\tau)$ from the learning period in Section 3. The testing period is all of the sample, January 1985 - December 2001. Therefore half of the results are in-sample and the rest are out-of-sample. In the next section we will use only out-of-sample results.

In Figure 5.1 we hedge a target asset with maturity of five years against a portfolio constructed of bonds with maturities of 1Y, 2Y, and 7Y. We use here only two factors for hedging. If we use three factors we get high weights for the bonds, because of the lack of constraints on the size of β. We solve $F+1$ linear equalities in order to construct the portfolios in Figures 5.1 through 5.4.
In the following figure, the x-axis represents the number of months passed, and the

y-axis is the cumulative excess return. The green line is the actual return and the blue line the theoretical return (drift).

Figure 5.1: Hedging 5y Vs. 1y2y7y

We present the cumulative practice and theoretical excess return as function of time, resulted from the hedging.



It is evident that the cumulative profit attained in practice is in excellent agreement with that predicted by the model, and the cumulative return is different from zero. The theoretical and actual variances are 1.26610^{-6} and 2.4910^{-6} respectively. The weights of each bond in the hedged portfolios over the entire period is shown in the next figure.

In order to show positive profit, we can see from Figure 5.2 that we need to change which bond we hedge against. Therefore we take the bond that is closest to maturity of two years and set its β to equal 1 (instead of the 5Y bond). Figure 5.3 shows portfolios that include the 2Y bond with $\beta = 1$ and three other bonds.

Figure 5.2: Weights of the portfolio over the entire period
We present the bonds weights we hold for the last portfolio.

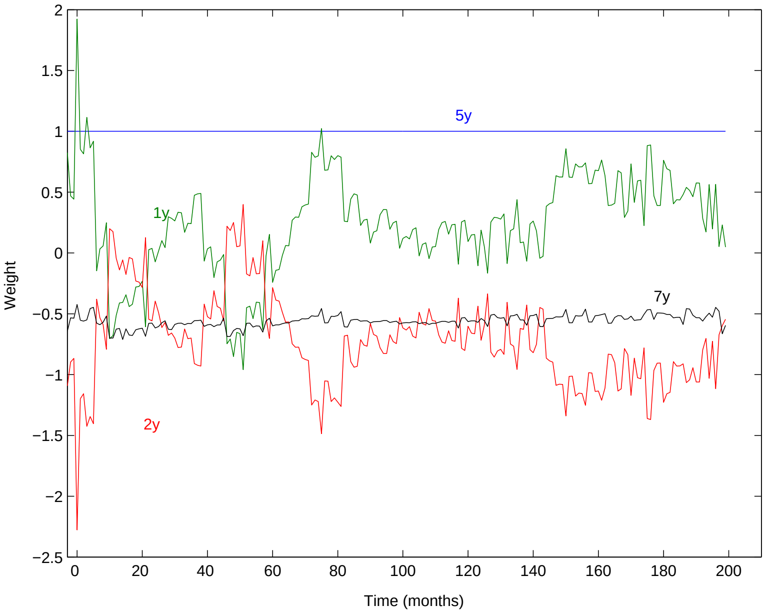
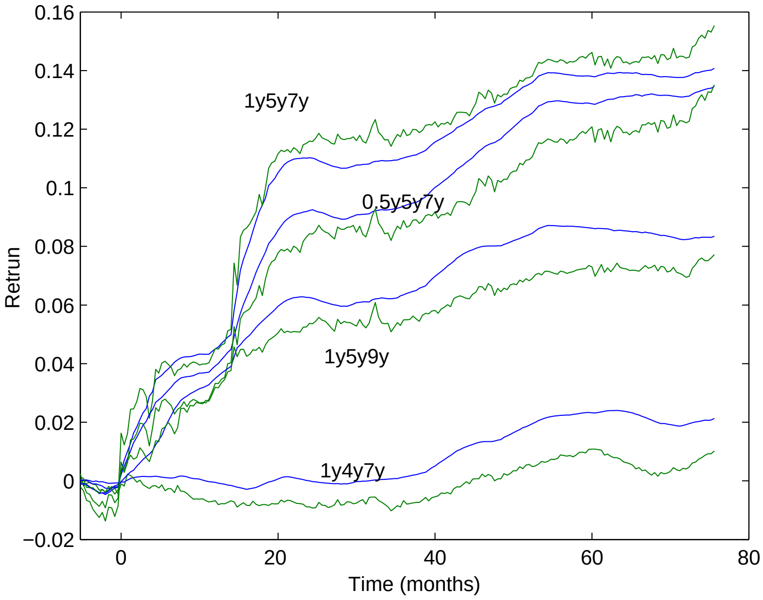


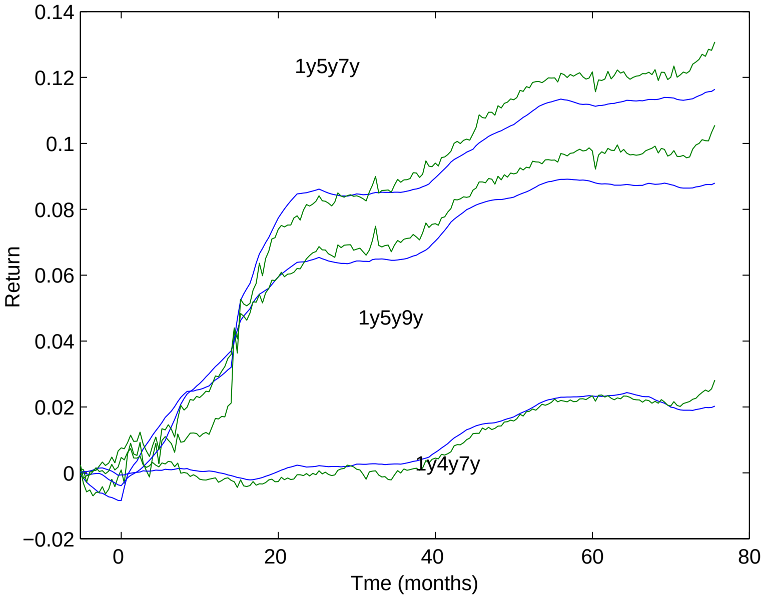
Figure 5.3: Hedging 2y vs. different bonds
We present the cumulative practice and theoretical excess return as function of time, resulted from the hedging.



In figure 5.3 a straightforward computation may provide a preliminary estimate of the magnitude of this phenomenon. Consider, as an example, the portfolio that consists of maturities 1Y, 5Y, 7Y and 2Y. The portfolio gained an actual return of 15.5 percent over 17 years, which is 0.92 percent per year.

To create cheaper schemas and to reduce the "noise" in the theoretical and actual returns, we will change the trading point. The new trading point will be every six months, meaning that every six months we will select new bonds with maturities closest to the defined ones. In the period between one to six months we will make minor adjustments in the βs of the chosen bonds every month, in order to maintain the three constraints.

Figure 5.4: Hedging 2y Vs. different bond-holding portfolios, 6 months
We present the cumulative practice and theoretical excess return as function of time, resulted from the hedging every six months.



Finally, in this section we use the theorem offered in Section 2 and offer simple examples for the excess returns in practice and in theory. The results show that Reisman and Zohar's (2004) model, when applied to real bonds, is an excellent predictor of the actual return. In the next section we will seek the optimal portfolio, so as to evaluate the magnitude of the theoretical return. The optimization will be performed instantaneously, so the optimal asset allocation will vary with time. We will also look for arbitrage opportunities in the bond market.

6 Advanced Applications

The purpose of this section is to show how to use the theorem presented above, in particular how eq. (2.8) can be exploited for applications. First we will show how to construct a portfolio with maximum expected returns using all available bonds. Second, to check the assumption that there are no arbitrage opportunities when transaction costs are included, a portfolio with net maximal returns will be constructed, meaning that at each trading point the new portfolio will include bonds with the maximum expected return subject to transaction costs. If the overall net return is positive and significantly different from zero, then there are arbitrage opportunities in the bond market.
The idea is to show that one can create arbitrage opportunities. In contrast to no-arbitrage models, which imply that all portfolios with the same risk exposures should yield the same expected return, the observation here is that in reality this is not the case. The existence of nonzero returns implies that such portfolios may yield different expected returns, and it is then desirable to pick a portfolio with the largest expected return.

The way to find the best zero-cost portfolio with no factor risk is to seek the one that maximizes the theoretical return. In this section the results are based on the modes suggested by Diebold *et al.* (2006). As before, we use those modes after applying the Gram-Schmidt procedure to them for orthonormalization purposes. Because these modes are used, the entire testing period is out-of-sample.

Excess Return Optimization

In this section we seek the best available portfolio at each trading point, meaning that at each trading point we will use all available bonds to find the maximum drift. By doing so we will reach the optimal cumulative return for the entire sample. We now add one more constraint and solve a maximization problem at each trading point. This means we will enter a new position every trading point and close the old one according to the problem's βs.

In order to solve the next problem we need at least $F+2$ bonds, and the target asset is zero. The maximization problem solved every period t is:

$\max_{\beta_j(t)} \sum_{j=1}^{M} \beta_j(t)\mu_j(t)$

$s.t$

$\sum_{j=1}^{M} \beta_j(t) D_{ij}(t) = 0, \forall i \in \{1,\ldots,F\}$

$\sum_{j=1}^{M} \beta_j(t) P_j(t,T) = 0$

$\sum_{j=1}^{M} \beta_j(t)^2 = 1$

The simplest solution is to use Lagrange multipliers. By solving this problem one can use all three factors without being concerned about the weights of each bond. It should be remembered that as here we are buying and selling many bonds, the transaction cost will go up, and this will be considered as well in the following paragraph.

The following three figures show the solution for the best available portfolio. The construction of the portfolio will be every month. The difference between the figures is in the number of factors used for hedging. In Figure 6.4 we hedge using one factor, a process similar to duration hedging. In Figure 6.5 the hedge uses two factors, and in Figure 6.6 three factors. We will also explore the optimal number of factors, using the Sharpe ratio criterion.

From Figure 6.1 we see that over almost seventeen years earnings stand at 40%, meaning a greater than two-percent riskless return for a year. The Sharpe ratio for this portfolio is 0.54 monthly or 1.83 annually.

In Figure 6.2 we again see an excellent fit between the theoretical and the actual lines. The actual line shows a return of almost 31%, which is the same as the prediction. The actual return for one year is more than 1.8%. The Sharpe ratio using two factors for hedging is 2 (annually), which is better than the Sharpe ratio using one factor.

Figure 6.1: Optimization Using One Factor for Hedging
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging using one factor.

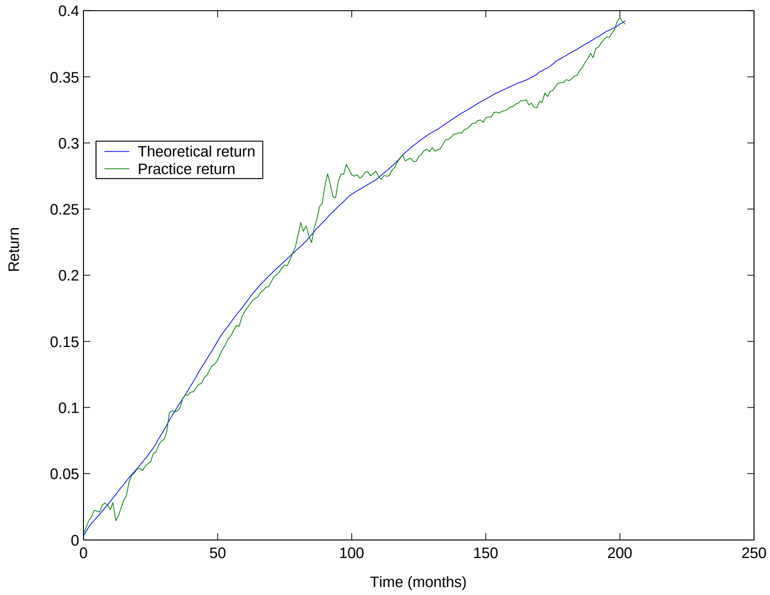
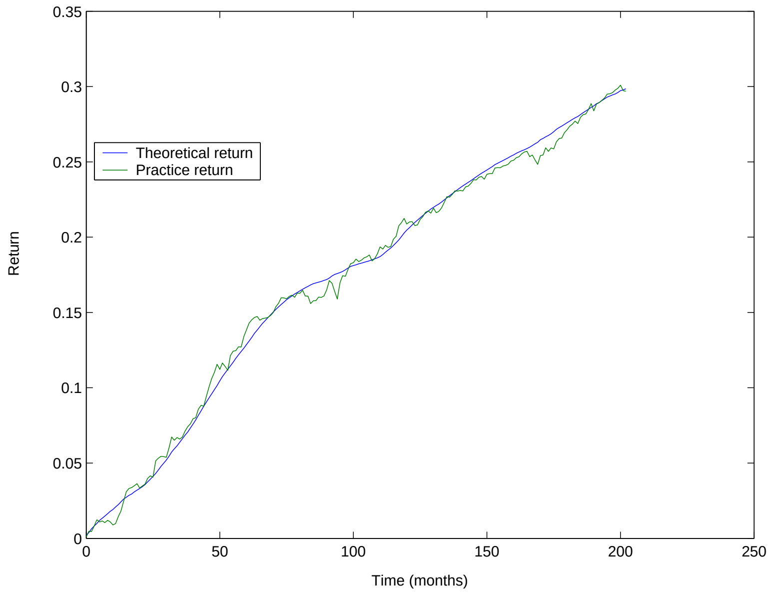
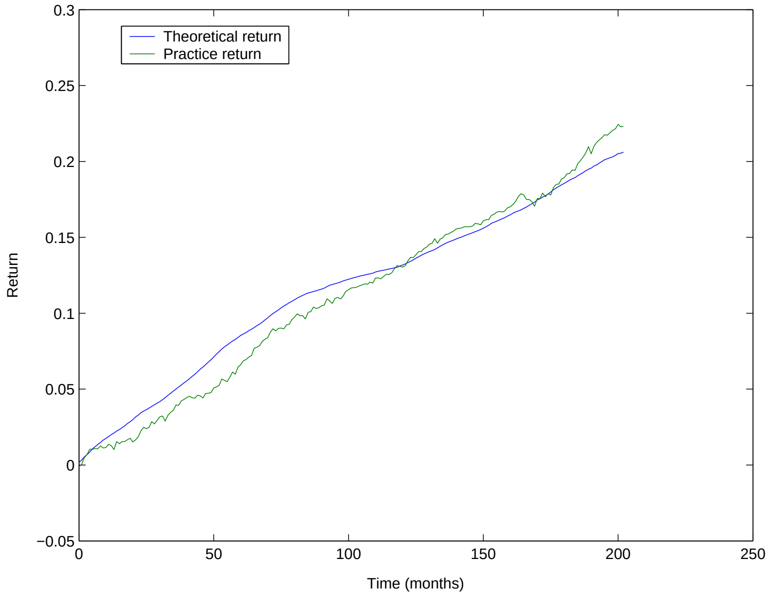


Figure 6.2: Optimization Using Two Factors for Hedging
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging using two factors.



This is because while the return is a bit smaller, the portfolio variance is reduced because the second factor captures an addition of more than seven percent of the term structure variation.

Figure 6.3: Optimization Using Three Factors for Hedging
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging using three factors.



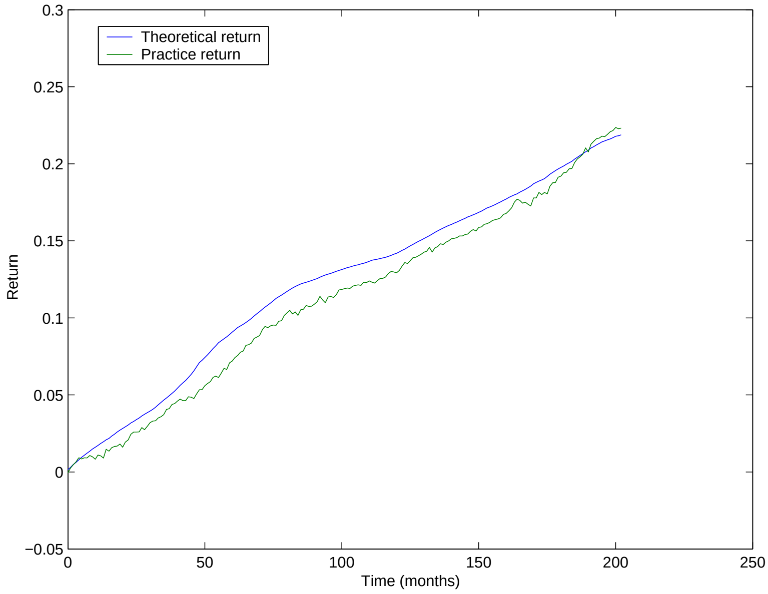
In Figure 6.3, we again see a good fit between the two lines. The actual return over seventeen years is 23% , versus a predicted return of over 20%. The computed Sharpe ratio is 0.63 for month or 2.147 annually, it is higher than the Sharpe ratio using two factors.

In summary, we learn that when selecting the best available bonds, in drift terms, we earn about one and a half percent return for a year (two factors and three factors). This is a riskless return. We showed also that the predicted return is similar to the return in practice. When comparing Sharpe ratios we see that using three factors gives the highest Sharpe ratio, and therefore the best performance. Using two factors gives a slightly lower Sharpe ratio. When there is only one factor (duration hedging), the Sharpe ratio is lowest. This is because when we use only one factor not all the risk is neutralized (there is still more than 10% unexplained risk in the term structure movement).

In Figure 6.4 we hedge using three factors, as before, but we hold the portfolio for six months. This will help bring down the "noise" caused by the buying and selling of bonds every month.

Figure 6.4: Optimization Using Three Factors for Hedging and Constructing a New Portfolio Every 6 Months

We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging done every six months using three factors.



We see an excellent fit, and a Sharpe ratio of 2.35 (annually). The theoretical cumulative return line and the actual line both reach almost 23%. The Sharpe ratio is the highest, because we have succeeded in bringing down the variance without the cost of losing returns.

Excess Return Optimization with Transaction Costs

It should be emphasized that while non-zero riskless profit may be obtained under ideal conditions (no transaction costs), in order to construct investment strategies that gain in practice one must consider individual trading costs. Constructing efficient investment strategies that take such considerations into account is our next interest. We will use almost the same maximization problem as in the previous section and add transaction costs. The solution is again via the Lagrange method, and includes all available bonds. One should note that the following optimization problem is not the ideal one. The idea is to show that even when we use a simple optimization to find the optimal return in the presence of transaction costs, arbitrage opportunities can be found. If the final actual net return is positive and high enough, the proposed theorem would be recommended in practice. Traders will be able to use more sophisticated methods for optimization to find larger net returns. If indeed we succeed in showing that there are net returns different from zero, we could argue that there are arbitrage opportunities in the bond market. The maximization problem solved every period t is :

$$\max_{\beta_j(t)} \sum_{j=1}^{M}[\beta_j(t)\mu_j(t) - TC(\beta_j(t) - \beta_j(t-1))^2]$$

$s.t$

$$\sum_{j=1}^{M} \beta_j(t) D_{ij}(t) = 0, \forall i \in \{1, \ldots, F\}$$

$$\sum_{j=1}^{M} \beta_j(t) P_j(t,T) = 0$$

$$\sum_{j=1}^{M} \beta_j(t)^2 = 1$$

where TC is a transaction costs that we choose. We use different values for this transaction costs: 1 cent and 5 cents for buying or selling one bond with a face value of 100. The following figures employ hedging using three factors, which produces the highest Sharpe ratio (as showed before). The construction of a new portfolio is set for every six months (with rebalancing every month). This will help to lower the transaction costs. As before, the blue line is the theoretical return and the green line the return in practice. The red line is the theoretical net return and the cyan line is the actual net return. Again the X-axis are months, and the Y-axis is cumulative return.

One can see that transaction costs of 1 cent for buying or selling a bond has very little effect on the returns. We see also that the actual net return is 18% and the Sharpe

Figure 6.5: Optimization with Transaction Costs of 1 cent for Bond
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging in the presence of transactions costs.

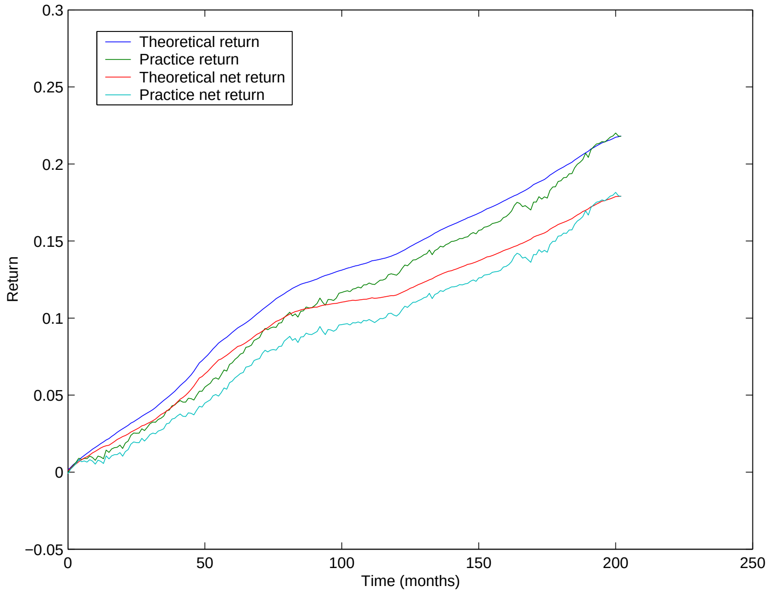
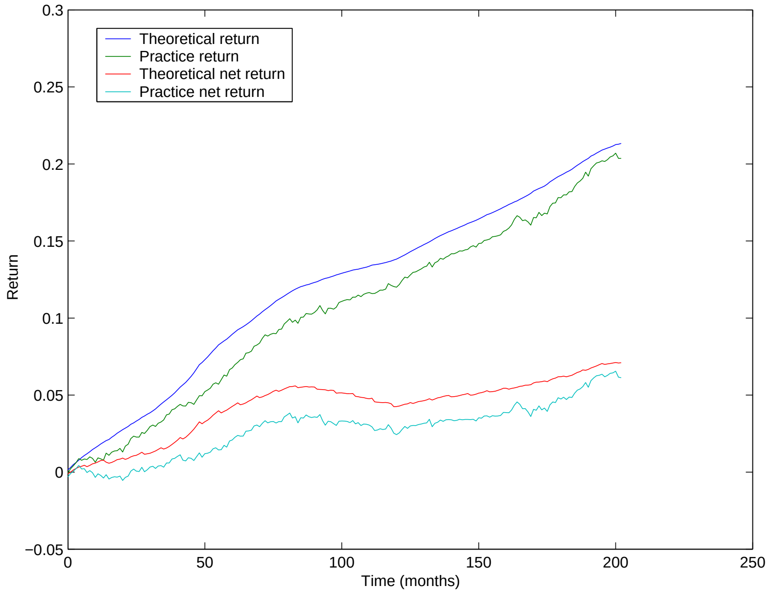


Figure 6.6: Optimization with Transaction Costs of 5 cents for Bond
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging in the presence of transactions costs.



ratio is 1.88. In the second of the two figures, we see that even when we use a 5-cent transaction costs we get a positive cumulative net return of around six percent and the Sharpe ratio is 0.65 (annually). It should be noted that constructing the portfolio every month gives similar results. When we use a transaction costs of 8-9 cents the overall net return is zero. Using transaction costs higher than that we start to lose money.

We have showed that when using an optimization with transaction costs, one can find positive net returns, meaning arbitrage opportunities. We used a simple optimization technique, suggesting that employing sophisticated optimization techniques will produce higher net returns.

The purpose of this chapter was to take the simple drift model and apply it for better use. Throughout this section we solved optimization problems based on seeking maximum theoretical returns under hedging constraints and a weights constraint. Using the last constraint, we were able to employ as many factors as we wanted.
In the first part we sought to construct the optimal bond portfolio, using all available bonds. We discovered that if we use three factors for hedging we get the best results in terms of combined return and variation. The optimized portfolio hedge using three factors gives an average excess riskless return of 1.35% for a year. Finally, we checked the arbitrage assumption, which states that there are no arbitrage opportunities when including transaction costs. We used a simple optimization technique, and found that up to a 7-cent transaction costs for buying or selling a bond does not preclude arbitrage opportunities.

7 Conclusions

We have presented the yield curve affine multi factor model. The multi-factor model specifies the spot interest rate as a linear combination of some F deterministic functions and stochastic coefficients. In this paper we used two methods in order to compute the deterministic functions (modes). First, we computed the modes using PCA as suggested by Litterman and Scheinkman (1991). In the second method we used Diebold, Li and Li's (2006) factor loadings.
We then used Reisman and Zohar's (2004) theorem for the next-period expected return and applied it to coupon-bearing bonds.

The model we suggest stands in contrast to existing models. We do not assume that there are no arbitrage opportunities in the market. Dai and Singleton (2000) showed that such an assumption might be destructive. We acknowledge the possibility that market frictions may allow the existence of what would have been arbitrage opportunities if the market had been frictionless. We therefore try to best fit the model to observed prices, and find that zero cost portfolios with no exposure to factor risk may indeed have returns that are substantially different from zero.

We first compared different risk management methods as in Diebold, Li and Li (2006b). We found similar results, showing that hedging with the multi-factor model outperforms the regular Macaulay duration model. Diebold, Li and Li (2006b) showed that the mean hedging error is different from zero. Here, using the drift formula, we explained most of this hedging error.
Second, we showed how to construct zero-cost zero-exposure portfolios. We found that the cumulative practice return is in excellent agreement with what the model predicts.
Finally, we searched for portfolios with maximal returns, and showed that the returns are positive and different from zero even if we take transaction costs into account.
We conclude that there are arbitrage opportunities in the bond market, and that the drift's formula is successful at predicting these opportunities.

The approach suggested here is very general, and can be applied to other derivatives markets as well. For example, the implied volatility smile can be described by a PCA model with a small number of factors. This model is not arbitrage-free, and so our methodology applies. It would also be interesting to determine why different bonds yield different expected returns and to characterize them by those returns.

8 APPENDICES

Appendix A - proof of eq. (2.6)

The price of bond at time t maturing at $T \equiv t + \tau$ is defined as $P(t,T) \equiv \sum_{k=1}^{K} c_i e^{-\tau_k y(t,\tau_k)} + 100 e^{-\tau_K y(t,\tau_K)}$ where T is the maturity and τ is time to maturity. The following theorem quantifies the drift $\mu(t,\tau)$ for the process $\frac{dP(t,T)}{P(t,T)}$.

Proof.

Given a term structure $y(t,\tau)$ which is an Ito process in t and for each fixed t is continuously differentiable in t,

Using Ito's Lemma,

$$dP(t,T) =$$

$$\sum_{k=1}^{K} \{ [(-\tfrac{\partial y(t,\tau_k)}{\partial t}\tau_k c_k + y(t,\tau_k)c_k)dt - \tau_k dy(t,\tau_k) + \tfrac{1}{2}(\tau_k)^2 d\langle y(t,\tau_k)\rangle] \exp(-y(t,\tau_k)\tau_k)\}$$

$$+100\exp(-y(t,\tau_K)\tau_K)[(-\tfrac{\partial y(t,\tau_K)}{\partial t}\tau_K + y(t,\tau_K))dt - \tau_K dy(t,\tau_k) + \tfrac{1}{2}(\tau_K)^2 d\langle y(t,\tau_K)\rangle] =$$

$$\mu_P(t) - \sum_{k=1}^{K} \{c_k \exp(-y(t,\tau_k)\tau_k) dy(t,\tau_k)\} - 100\exp(-y(t,\tau_k)\tau_k)\tau_K dy(t,\tau_k), \quad (8.1)$$

with $\mu_P(t)$ as in eq. (2.8) and $dy(t,\tau_k) = \sum_{i=1}^{F} D_i(t) dz_i(t)$. $D_i(t)$ defined in eq. (2.7).

Appendix B - Term Structure

We present three additional term structures with the exact dates as presented by Diebold and Li (2006). The X-axis is the maturity in years, and the Y-axis is the yield. As before, the term structure is the green line. The bid rates are the cyan points, and the ask rates are the red points.

Figure 8.1: Term Structure March 1989

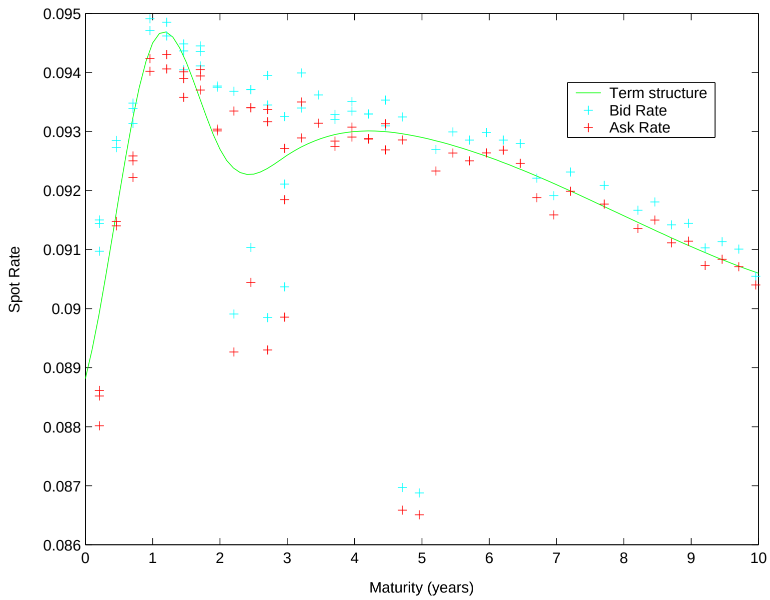


Figure 8.2: Term Structure July 1989

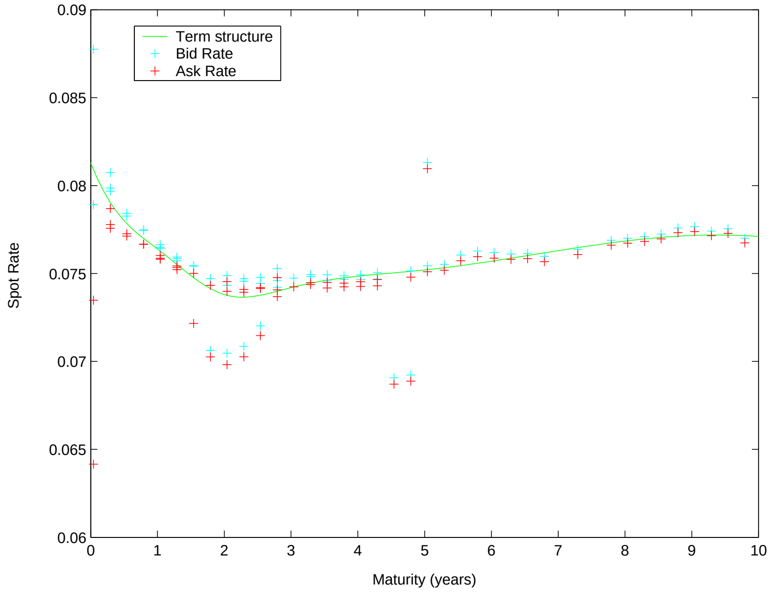
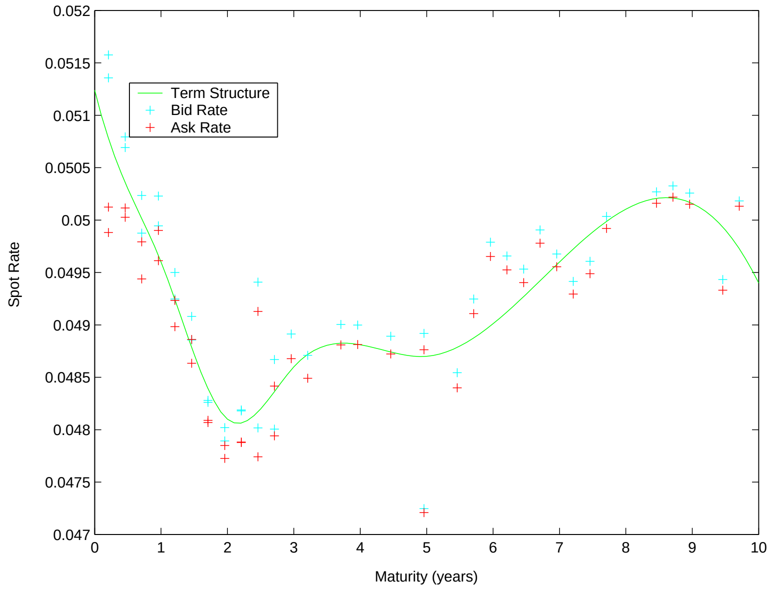


Figure 8.3: Term Structure August 1998

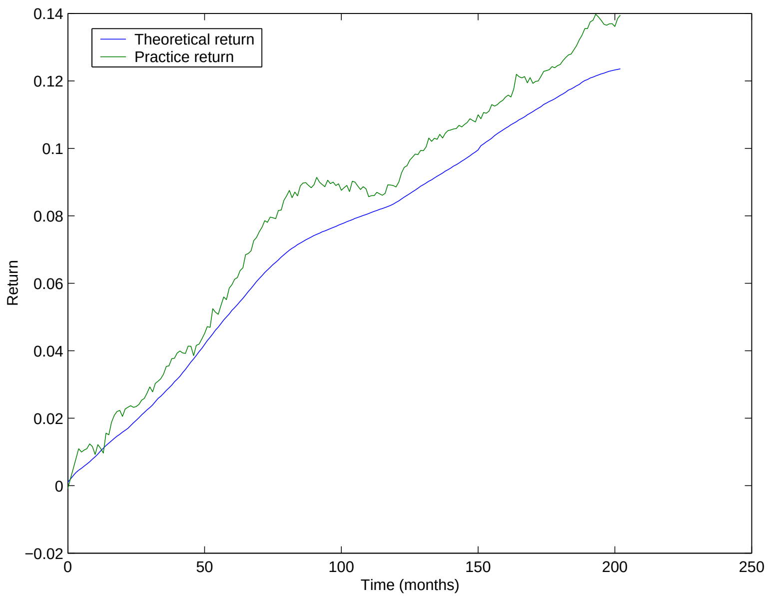


Appendix C - Advanced Application using new Term Structure

Here we will demonstrate the excellent predictability of eq. (2.8), using a simpler term structure estimation procedure. The term structure is estimated by the second method we mentioned, meaning we estimate the price using eq. (2.2) and search for the minimum sum square of price errors. Our dependent variables are again z_i. As modes we use Diebold, Li and Li's (2006b) factor loadings. In order to be accurate at the short end of the term structure, we multiply each price error by a weight of $1/T$, where T is the maturity of a specific bond. This estimation procedure is more familiar than the one we used.

We next present the same figures from excess returns optimization. In figure 8.4 we use three factors for hedging (Diebold *et al.* (2006) modes), and a holding period of 1 month. On the X-axis we put months and on Y-axis we put cumulative return. Again we find excellent agreement between the two lines, and the Sharpe ratio is 1.7.

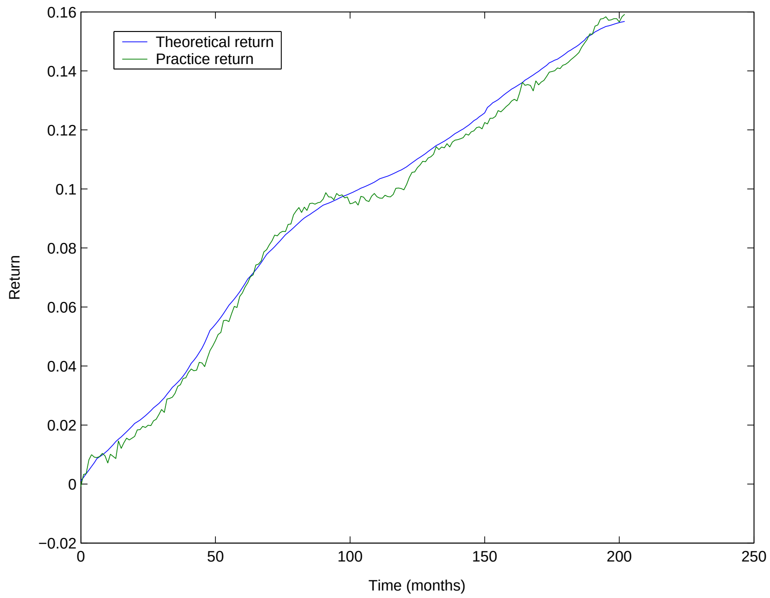
Figure 8.4: Optimization Using Three Factors for Hedging
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging.



In Figure 8.5 we use three factors, but a holding period of six months. Here the fit between the predicted and actual return is even better and the Sharpe ratio is 1.9.

Figure 8.5: Optimization Using Three Factors for Hedging and Constructing New Portfolio Every 6 Month

We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging done every six months.



We conclude that Reisman and Zohar's (2004) drift theorem is not dependent on how we estimate the term structure.

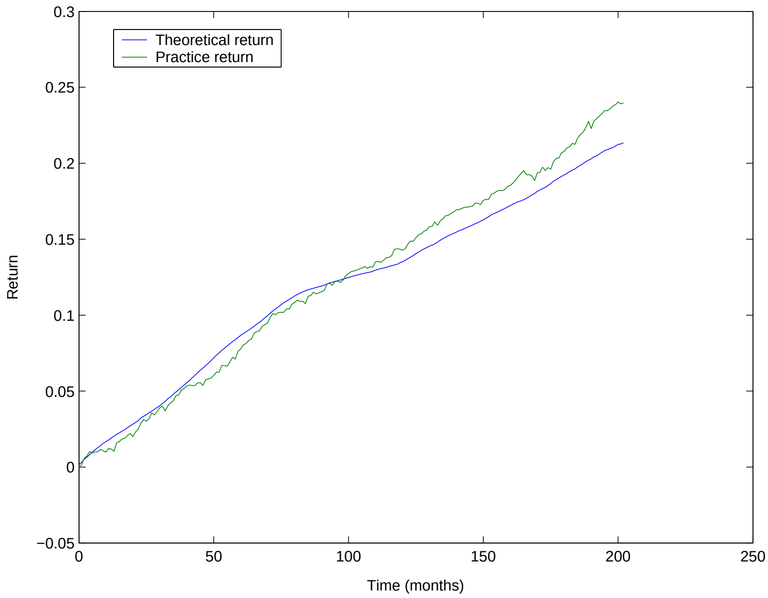
Appendix D - Advanced Application with Different Values for λ

We now want to present the cumulative actual and predicted returns using different values for λ.

First we present the same figures as 6.6 and 6.7, meaning optimization on all available bonds using three factors with holding periods of one and six months. In figures 8.6 and 8.7 we present the λ that brings to maximum the third factor loading with a maturity of three years. In Figure 8.6 we see an excellent fit between the returns,

Figure 8.6: Optimization Using Three Factors for Hedging

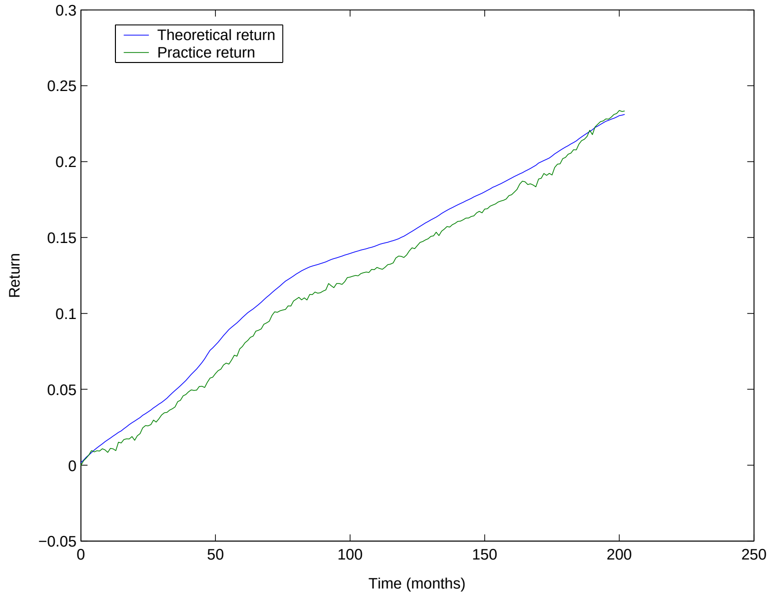
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging. $\lambda = 0.786$.



and the actual return is 24%. The Sharpe ratio is 2.32. In Figure 8.7 the actual return is 23.4% and the Sharpe ratio is 2.55.

Figure 8.7: Optimization Using Three Factors for Hedging Constructing New Portfolio Every 6 Months

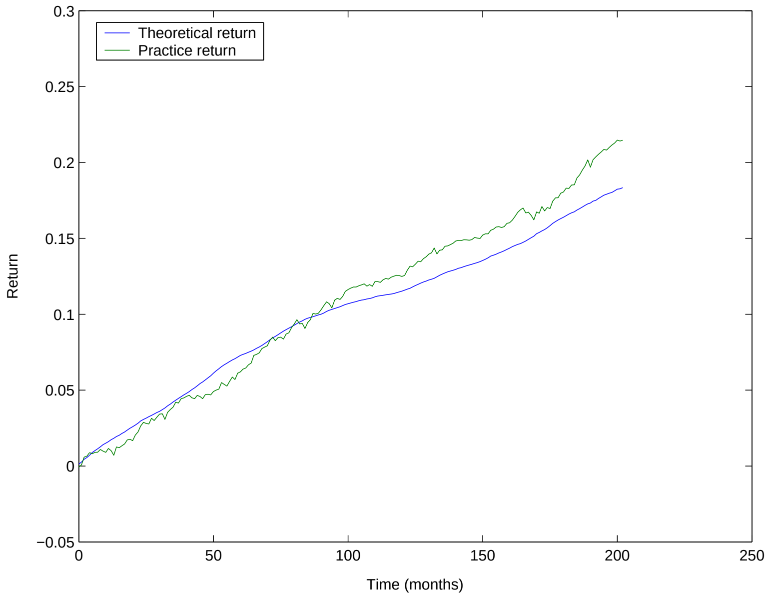
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging done every six months. $\lambda = 0.786$.



In Figures 8.8 and 8.9 we present the same schemes, but with λ that brings to maximum the third loading with a maturity of two years.

Figure 8.8: Optimization Using Three Factors for Hedging

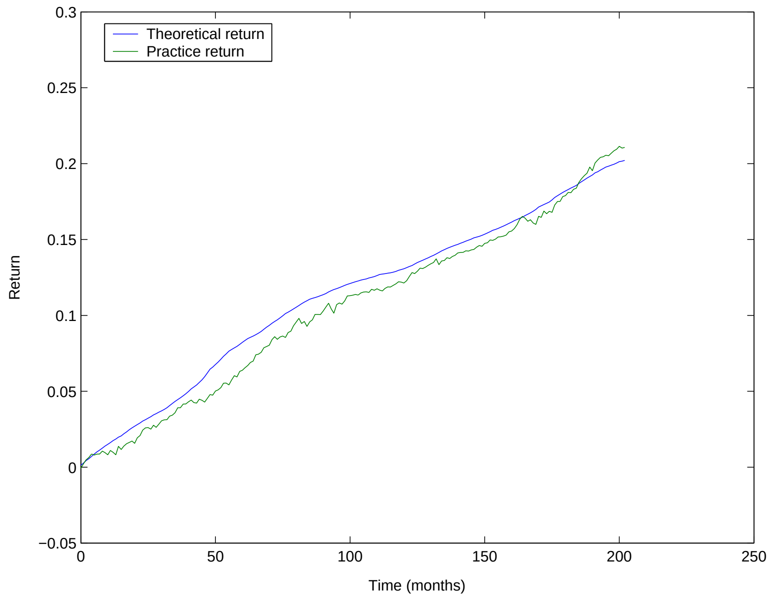
We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging done every six months. $\lambda = 0.584$.



In Figures 8.8 and 8.9 we again see excellent fit between the returns. The actual return in Figure 8.8 is 21.4% and the Sharpe ratio is 1.9. In figure 8.9 the actual return is 21.1% and the Sharpe ratio is 2.08.

Figure 8.9: Optimization Using Three Factors for Hedging Constructing New Portfolio Every 6 Months

We present the cumulative practice and theoretical excess return as function of time, resulted from the optimized hedging done every six months. $\lambda = 0.584$.



We conclude that the λ affects both the actual and predicted returns, but not by much. We can also see that as the λ gets smaller, the returns are higher (when $\lambda = 0.0609$ the return is 27%).

References

[1] Bierwag, G.O., G.G. Kaufman and A. Toevs (1983) "Duration: Its Development and Use in Bond Portfolio Management", *Financial Analysts Journal 39*, 15-35.

[2] Bliss R. and E.F. Fama (1987) "The information in long maturity forward rates", *American Economic Review 77*, 680 -692.

[3] Chambers, D.R., W.T Li and R.W. McEnally (1988) "Immunizing default-free bond portfolios with a duration vector", *The Journal of Financial and Quantitative Analysis 23*, 89-104.

[4] Dai, Q. and K. Singleton (2000), "Specification analysis of affine term structure models", *Journal of Finance LV(5)*, 1943-1978.

[5] Diebold, F.X. and C. Li (2006) "Forecasting the term structure of government bond yields", *Journal of Econometrics 130*, 337-364.

[6] Diebold, F.X., J. Li and C. Li (2006b), "A Three - Factor Yield Curve Model: Non - Affine structure, systematic risk sources, and Generalized", *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando*, 240-274 Cheltenham, U.K.: Edward Elgar.

[7] Duffie, D and R. Kan (1996), "A yield-factor model of interest rates", *Mathematical Finance 6 (4)*, 379-406.

[8] Gultekin, N.B and R.J Rogalski (1984), "Alternative Duration Specifications and the Measurement of Basis Risk: Empirical Tests", *The Journal of Business 57*, 241-264.

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

[10] Nelson, C.R. and A.F. Siegel (1987), "Parsimonious Modeling of Yield Curves", *Journal of Business 60*, 473-489.

[11] Reisman, H. and G. Zohar (2004), "Excess yields in bond hedging", *RISK 17 (12)*, 32-39.

[12] Reisman, H. and G. Zohar (2004b), “Instantaneous mean-variance analysis of bond returns”, *Journal of Fixed Income 14 (1)*, 32-39.

[13] Reisman, H. and G. Zohar (2004c), “Short Term Predctability of The Term Structure”, *Journal of Fixed Income 11 (1)*, 7-14.