# Excess yields in bond hedging

Haim Reisman and Gady Zohar

Department of IE & Management

Technion - Israel Institute of Technology

Haifa 32000, Israel

reisman@ie.technion.ac.il ; gady@tx.technion.ac.il

April 12, 2004

**Abstract**

We explore a dynamic term structure factor model that implicitly allows for arbitrage opportunities, and we estimate it on Treasury data. Using this model we construct instantaneously risk free portfolios, and we write down a formula for their possibly non-zero excess returns. Our model anticipates that such excess yields may be quite large. When testing the performance of these portfolios we find that their returns in practice perfectly match what our model predicts. An important implication of our approach is that hedging against factor risk may involve substantial excess gains or losses, that can be determined by our model.

*Key words: Term Structure, Principal Component Analysis, Dynamic Hedging, Convenience Yield.*

# 1 Introduction

In this paper we suggest a dynamic term structure model that explicitly allows for arbitrage opportunities, and estimate the model on Treasury data to determine the magnitude of these opportunities. Of course, arbitrage opportunities do not really exist; the interpretation of the apparent arbitrage opportunities is that market frictions prevent investors from exploiting them.

We suggest this model in contrast to the usual approach of imposing no-arbitrage in dynamic term structure models (as in HJM (1992), Duffie and Kan (1996), and others). In the construction of no-arbitrage models one makes the strong assumption that market frictions do not exist, while deviations from no-arbitrage are typically interpreted as noise. However, markets are not without frictions, and therefore a model that explicitly allows for arbitrage opportunities may be more useful than the standard approach if it can tell something about the magnitude of these opportunities.

A model that is not arbitrage-free has been suggested by Litterman and Scheinkman (1991). They show that the term structure of interest rates is reliably modeled by an affine three factor model attained by principal component analysis (PCA). However, they disregard the possible deviations from no-arbitrage, presuming these are small.

In this work we adopt the factor model of Litterman and Scheinkman (1991), and model its deviations from no-arbitrage. We construct zero cost portfolios that have no exposure to factor risk, and are therefore instantaneously risk free. While such portfolios should have no return in an arbitrage free model, in our model the return on such a portfolio may be non-zero, and we call it *convenience yield*.

We derive an analytical formula for the convenience yield and estimate the parameters of the formula using historical data of US bond markets. We may use the formula to theoretically determine the instantaneous convenience yield of any zero cost portfolio with zero instantaneous variance. For the same portfolio we then measure the return that is attained in practice. By comparing the long-run performance of the portfolio with our theoretic projections we find that our formula determines the convenience yield with striking accuracy, both in sample and out of sample. Furthermore, the convenience yields of zero cost portfolios with instantaneous zero variance may be quite large: we easily construct such portfolios with yearly return of almost one percent of the long (or short) side of the portfolio.

Different kinds of frictions may cause such large convenience yields. For example, repo specialness can make certain bonds more attractive than others. However, it is important to emphasize that in this work we do not attempt to determine what may be the reasons for excess yields to exist, but only to quantify such yields when they exist.

The most straightforward application of this result is in hedging. When standard no-arbitrage models are used for hedging, the hedged portfolio will always return the risk free rate. This is not the case when our model is used, since the hedging results with an excess yield that depend on the specific instruments that we use for the hedging. Surprising as it may sound, the empirical results strongly support our model: indeed when we hedge a certain bond using our model, we get different excess yields that depend on the specific bonds used for the hedging, and these excess yields perfectly agree with our theoretic prediction.

Closely related to the methodology presented in this paper is the (potential) predictability of the term structure. Several recent works address this issue (see Diebold and Li (2002) and references therein). Indeed, if changes in the term structure are predictable, one will make excess gains by holding an appropriate portfolio. Dai and Singleton (2003) show that certain no-arbitrage models are capable of accounting for such predictability.

Nonetheless, our methodology is fundamentally different. While forecasting is actually betting on the direction in which the leading factors of uncertainty change, our incentive is to actually construct portfolios with no exposure to these factors at all. Even if one can predict future changes in the factors, these predictions are completely ineffectual in a portfolio with no exposure to these factors!

A remarkable feature of our work is that the model works so well despite the use of rather elementary specification and estimation method. It turns out that a specification of the actual term structure dynamics is not needed in order to determine the excess return of portfolios with zero instantaneous variance.

In order to demonstrate the robustness of our model, we also apply our formula to simulated data that is generated by the Vasicek model. First we synthetically generate prices that contain no arbitrage opportunities, and as expected we find zero convenience-yield. Then we synthetically add arbitrage to the prices, and use our formula to find a portfolio with significant non-zero convenience yield. The yields attained in practice (using the synthetically generated prices) exactly match those predicted by the model.

The paper is organized as follows. In section 2 we specify the model, show how risk free portfolios should be constructed within our framework, and derive a formula for the convenience yield of such portfolios. We also show how to find a risk free portfolio for which the convenience yield is maximal.

In section 3 we test our approach on simulated data. First we synthetically generate bond prices that contain no arbitrage opportunities, and as expected we fail to find a statistically significant non-zero convenience yield. Then we perturb the prices, and use our formula to indeed recover non vanishing convenience yield. Hence we are able to quantify the excess return that is attained as a result of a small perturbation in price. This simulation demonstrates the robustness of our approach despite the use of relatively simple interpolation and estimation methods.

In section 4 we focus on US Treasuries. We estimate the model parameters and compute the theoretical time dependent convenience yields in hedging certain bonds using different sets of bonds. The yearly convenience yield in some cases is as high as 0.4% of the long side of the position. The agreement between the theory and practice is striking. We then compute the zero cost zero risk portfolio that maximizes the convenience yield under some norm constraint. In this case the yearly convenience yield can be as high as 1% of the long side of the position and the Sharpe ratio of the convenience yield in this case exceeds 2 in annual terms.

While so far we discussed only convenience yields of portfolios with zero instantaneous variance, in section 5 we suggest a way to define the convenience yield of specific bonds so that all convenience yields are zero iff there is no arbitrage.

We conclude in section 6 with a short discussion.

## 2 The convenience yield

In this section we specify our model for a market with $N$ discount bonds[1] of maturities $T_1, \ldots, T_N$. We derive an expression for the drift of individual bonds, and fit a factor model to account for the randomness in the yields. We then show how to construct zero cost portfolios with zero exposure to the leading factors of uncertainty. Finally, we write down the formula for the excess yield of such portfolios, which we define as the convenience yield.

Denote by $y(t,\tau)$ the spot rate for $\tau$ years at time $t$. In this paper we use $T$ and $\tau$ interchangeably to denote maturity and time to maturity of the same bond, namely, $T \equiv t + \tau$. Similarly, for a bond indexed by $i \in \{1, \ldots, N\}$ we identify $T_i \equiv t + \tau_i$.

$B(t,T) \equiv e^{-\tau y(t,\tau)}$ is the price at time $t$ of a discount bond maturing at $T \equiv t + \tau$. Given \$1 invested in a discount bond, we construct a zero cost portfolio by borrowing \$1 at the instantaneous risk free rate, $r(t)$. The instantaneous change in the value of such a portfolio is $dB(t,T)/B(t,T) - r(t)dt$. The following theorem (proven in appendix B) quantifies the drift ($\mu(t,\tau)$) of these processes:

**Theorem 2.1** *Given a dynamically evolving term structure $y(t,\tau)$ which is differentiable in $\tau$,*

$$\frac{dB(t,T)}{B(t,T)} - r(t)dt = \mu(t,\tau)dt - \tau dy(t,\tau), \tag{2.1}$$

*where*

$$\mu(t,\tau) := y(t,\tau) - r(t) + \tau\frac{\partial y(t,\tau)}{\partial \tau} + \frac{1}{2}\tau^2\frac{d\langle y(t,\tau)\rangle}{dt}, \tag{2.2}$$

*and $\langle\cdot\rangle$ denotes the quadratic variation.*

$\partial y(t,\tau)/\partial\tau$ is the slope of the term structure at $\tau$. Given $N$ spot rates, in order to compute this slope we must interpolate the term structure. For the results presented

[1]The theory we develop can be fitted also for a continuum of maturities in a straightforward manner.

in this paper we use cubic spline. However, when we test with different interpolation techniques the results that we obtain are very similar.

The estimation of the quadratic variation term $d\langle y(t,\tau)\rangle$ in eq. (2.2) should be based on the observed variation. With continuous-time data this quantity can be estimated with infinite accuracy. In this work we use monthly data and maturities up to 10 years, and then it turns out that the quadratic variation term hardly affects the results.

While Theorem 2.1 completely specifies the drift $\mu(t,\tau)$, a model is needed in order to specify $dy(t,\tau)$. We assume that the economy is driven by $F$ factors of uncertainty $z_i$, $i = 1,\ldots,F$, and that the spot rate for $\tau$ years admits the specification $y(t,\tau) = \sum_{i=1}^{F} u_i(\tau)z_i(t) + \epsilon(t,\tau)$, which immediately implies that

$$dy(t,\tau) = \sum_{i=1}^{F} u_i(\tau)dz_i(t) + d\epsilon(t,\tau). \tag{2.3}$$

One may choose $u_i(\cdot)$ arbitrarily, but such a representation becomes useful only if the residual noise $d\epsilon$ is small. Consequently we set $u_i(\tau)$ as the (orthonormal) principal directions determined by applying standard principal component analysis (PCA) to a sequence of vectors $(dy(t,\tau_1),\ldots,dy(t,\tau_N)) \in \mathbb{R}^N$ (for discrete times $t$)[2]. The procedure that we use is portrayed in detail in Appendix A.

By taking $u_i$ as the principal directions, and then defining $dz_i$, $i = 1,\ldots,F$ as the orthonormal projections of $dy(t,\cdot)$ onto $u_i(\cdot)$,

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

we minimize the in-sample mean squared error $\sum_t \sum_{i=1}^{N} d\epsilon^2(t,\tau_i)$ (Jolliffe (1986)).

Since (2.3) is the only assumption we make on $dy$, we can say nothing regarding

[2] There are many other reasonable choices of $u_i$. One example is the Nelson-Siegel (1987) factor loadings.

the statistical properties of neither $z_i(t)$, $i \in \{1, \dots, F\}$, nor $\epsilon(t, \tau)$. If the term structure evolves in a complex manner, so will the factors $z_i(t)$. In particular, these are not assumed to be martingales, and may have non-zero correlations. Nonetheless, for a given $F$, the obtained factors will explain most of the in-sample variance in the sense of minimizing the squared error above.

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 that it requires no ad-hoc assumptions on the dynamics of processes. No assumptions are needed for the PCA to recover the most significant principal directions *in sample*, but this does not guarantee that the principal vectors will perform well *out of sample*. Stationarity is a sufficient (but not a necessary) condition for the PCA to perform well also on out of sample data.

The number of factors, $F$, should be as small as possible, yet large enough to make the residual noise $d\epsilon(t, \tau)$ negligible. In practice, strong correlations among bonds of different maturities lead to reliable characterization of the dynamics of all bond prices with $F$ significantly smaller than $N$. Typically, about 99% of the variance in bond yields can be explained by as few as three factors.

For the rest of this paper denote a bond portfolio at time $t$ by

$$\pi(t) \equiv (\pi_1(t), \dots, \pi_N(t))^T,$$

where $\pi_i(t)$ $(i = 1, \dots, N)$ is the amount of money at time $t$ in bond of maturity $T_i$.

**Constructing portfolios with no factor risk.** To this end we concentrate on zero cost portfolios with no factor risk. Denote the value of a given portfolio $\pi$ by: $V \equiv \sum_i \pi_i$. The instantaneous change in $V$ is:

$$dV = \sum_{i=1}^{N} \pi_i(t) \frac{dB(t,T_i)}{B(t,T_i)}. \tag{2.4}$$

(2.1) with (2.3) imply that the *exposure* of this portfolio to factor $k$ is: $\delta_k \equiv \sum_{i=1}^{N} \pi_i \tau_i u_k(\tau_i)$. Hence, the value of any portfolio $\pi(t)$ is independent of $dz_k$ iff

$$\delta_k \equiv \sum_{i=1}^{N} \pi_i(t) \tau_i u_k(\tau_i) = 0. \tag{2.5}$$

Any portfolio that satisfies (2.5) for all $k \in \{1, \ldots, F\}$ entails no factor risk.

Then it takes at least $(F+2)$ bonds to construct a zero cost riskless portfolio. Given $(F+2)$ bonds, such a portfolio is simply a solution of a set of linear equations which is detailed in appendix C. Note that these portfolios consist of constant weights since we assume that the $u_i(\tau)$ do not vary in time.

With the assumption that there is no residual risk, (2.1), (2.3) and (2.4) imply that the excess return of a zero cost portfolio $\pi$ with no exposure to factor risk is:

$$\mu(\pi) := \sum_{i=1}^{N} \pi_i(t) \mu(t, \tau_i). \tag{2.6}$$

**Definition.** The *convenience yield* at time $t$ of a zero cost portfolio $\pi(t)$ with no exposure to factor risk is its excess return, $\mu(\pi(t))$.

The formula (2.6) for the convenience yield is accurate up to an error term that is bounded by the mean squared residual error that was mentioned earlier. To see this, note that the instantaneous return on the portfolio $\pi(t)$ is actually

$\sum_{i=1}^{N} \pi_i(t)(\mu(t,\tau_i)dt + d\epsilon(t,\tau_i))$. It follows from the Cauchy-Schwartz inequality that the deviation of this return from the one implied by the theoretical convenience yield is therefore bounded by

$$|\sum_{i=1}^{N} \pi_i(t) d\epsilon(t,\tau)| \leq \|\pi(t)\| \|d\epsilon(t,\cdot)\|.$$

For a zero cost dynamic allocation with no factor risk and norm of unity, the error in the cumulative profit is therefore bounded by $\sum_t \sum_{i=1}^{N} d\epsilon^2(t,\tau_i)$. We saw that this quantity is minimized (in sample) by using PCA, and that is an additional important justification for using PCA.

Since the number of leading factors is smaller than the number of traded securities, there is a continuum of zero cost instantaneously risk free portfolios. In order to evaluate the magnitude of the convenience yield we would like to find the zero cost risk free portfolio which maximize the convenience yield. A formulation of such an optimization problem and its solution are presented in appendix C. Since the optimization is performed instantaneously, the optimal asset allocation varies in time.

# 3 Testing in simulation

In this section we test our methodology in simulation. We assume Vasicek dynamics for the term structure of interest rates. When yields follow this simple model, the arbitrage free price dynamics is well known (see for example Hull (1997), page 419). Hence, for all maturities we synthetically generate weekly prices that are arbitrage free with respect to this model. We then estimate the factor loadings in our model, find the convenience yields for all bonds, find the optimal dynamic allocation that will produce maximal gains, and find that indeed there is no arbitrage.

We begin by assuming that the instantaneous risk free rate is modeled by the following stochastic process:

$$dr = a(b - r)dt + \sigma dW, \tag{3.1}$$

where we use as an example $a = 0.05$, $b = 0.05$, $\sigma = 0.02$, $r_0 = 7\%$. We iterate this equation on a weekly basis, pick a realization in which rates are never negative, and determine the prices of discount bonds maturing at $3M$, $6M$, $1Y$, $2Y$, $3Y$, $5Y$, $7Y$ and $10Y$ according to the equation (Hull (1997), page 419):

$$B(t, T) = \alpha(t, T) \exp\left(-\beta(t, T) r(t)\right), \tag{3.2}$$

where

$$\beta(t, T) = \frac{1 - exp(-a(T - t))}{a},$$

$$\alpha(t, T) = \exp\left[\frac{(\beta(t, T) - T + t)(a^2 b - \sigma^2/2)}{\sigma^2} - \frac{\sigma^2 \beta^2(t, T)}{4a}\right].$$

Based on these prices as raw data we first use PCA to determine the leading factors. Not surprisingly we find that the first factor accounts for 100% of the variation in the term structure.

In Figure 1 we plot the cumulative (out-of-sample) profit from trading optimally in all bonds according to the scheme suggested in section 2, while using the prices attained from the simulation. The norm of the portfolio is set to $100 while the portfolio has a net worth of zero at all times. Gains are not reinvested so the slope of this plot is the instantaneous rate of profit from the portfolio while restarting every week from a normalized zero cost position. A dashed line stands for our theoretical prediction while solid lines show profit obtained by trading in practice.

As is evident from Figure 1, our model arrives at the conclusion that the convenience yield is zero, and arbitrage gains cannot be made. Moreover, the profit attained from trading optimally in the simulated prices is in excellent agreement with the theoretical prediction.

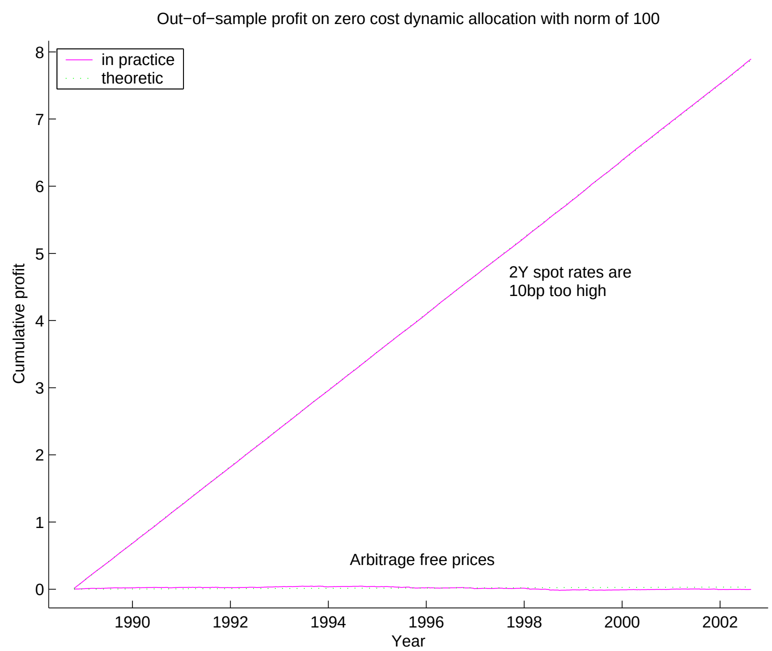



Figure 1: Profit from zero cost zero variance portfolios on prices generated using Vasicek model

To generate prices that are inconsistent with arbitrage we set the price of the 2Y discount bond lower so that its yield to maturity is 10 basis points higher than the yield implied by the arbitrage free model. This means that the term structure has a tiny bump around $T = 2Y$, that does not vanish in time. This clearly generates an arbitrage opportunity, and this opportunity is fully exploited by our optimized procedure: we now get an upward sloping cumulative profit plot where the slope, which is the instantaneous rate of profit, is maximized. Once again, the theoretical profit and the profit in practice are almost identical.

Despite the small mispricing in the single perturbed bond, the convenience yield that can be attained by a properly chosen portfolio is quite large. In this portfolio both long and short legs worth about \$125, which means that the convenience yield is approximately 0.5% of each leg.

The amount invested in each bond is plotted in figure 2 (the optimal portfolio in this case hardly varies in time). The consistent bump was clearly detected by our scheme since the emerging optimal strategy sells short 3Y bonds and buys 2Y bonds, both producing excess capital gains due to the 2Y excessive YTM. The rest of the weights though are not zero in order to make the portfolio zero cost and without exposure to risk.

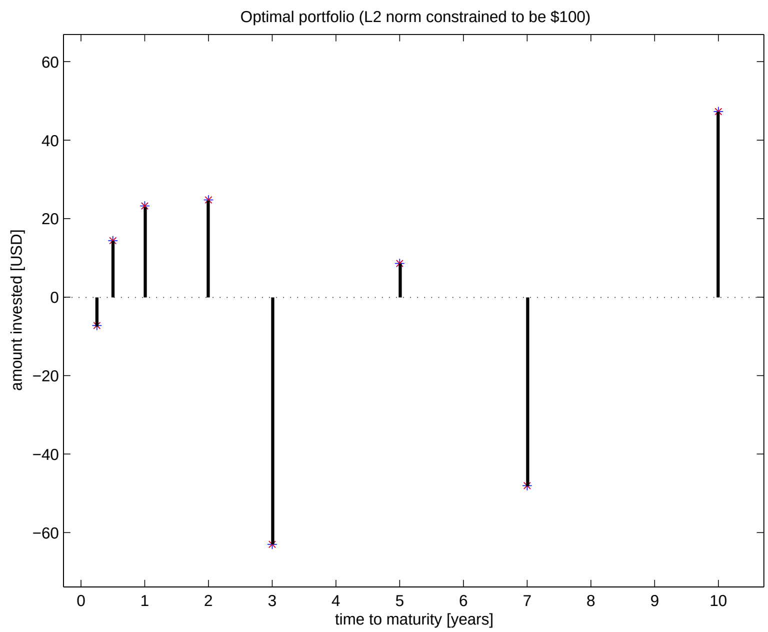



Figure 2: Optimal portfolio when prices from Vasicek model are perturbed

This simulation clarifies the strength of our methodology: first we show how to choose a risk free portfolio that captures imperfections in term structure dynamics and exploits them by maximizing the risk free gains. Additionally, our approach suggests these excess gains (0.5% in this example) as a measure of how far is the observed term structure dynamics from being consistent with no-arbitrage. Finally, we accurately predict what this excess yield is expected to be using formula (2.2).

# 4 Convenience yields in US bonds

In this section we apply our methodology to the US bond market. We start by estimating the leading principal directions $u_i(\tau)$, and for ant $t$ we compute the drifts $\mu(t, \tau_i)$ for all maturities $\tau_i$. While hedging single bonds using other sets of bonds we demonstrate the existence of non-zero convenience yields which may vary substantially as a function of the assets used for hedging. Finally we find a dynamic zero cost allocation with zero instantaneous variance which maximizes the excess return. For each portfolio we compare the expected rate of return with the returns obtained in practice in order to demonstrate the good fit between the two.

Our data includes spot rates of US Treasuries for 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, and 10Y. These numbers are provided by the US federal reserve on a weekly basis. We use the data between June 1987 till June 2002. Even though in-sample testing produces more impressive results, we focus on out-of-sample testing, since this is more interesting from a practical point of view. This means that we determine the principal directions $u_i(\tau)$ based on the first half of the data set (1987:6-1994:12), and test on the second half (1995:1-2002:6). Since the data is weekly, we approximate monthly data by considering each fourth measurement.

**The principal directions.** Based on our data set we apply PCA (as detailed in Appendix A) on the vectors $(dy(t, \tau_1), \ldots, dy(t, \tau_N))$ for all $t$ in the learning period, and determine the principal directions $u_i(\cdot)$. The three leading principal directions that are based on the learning period (1987-1994) are drawn in Figure 3.

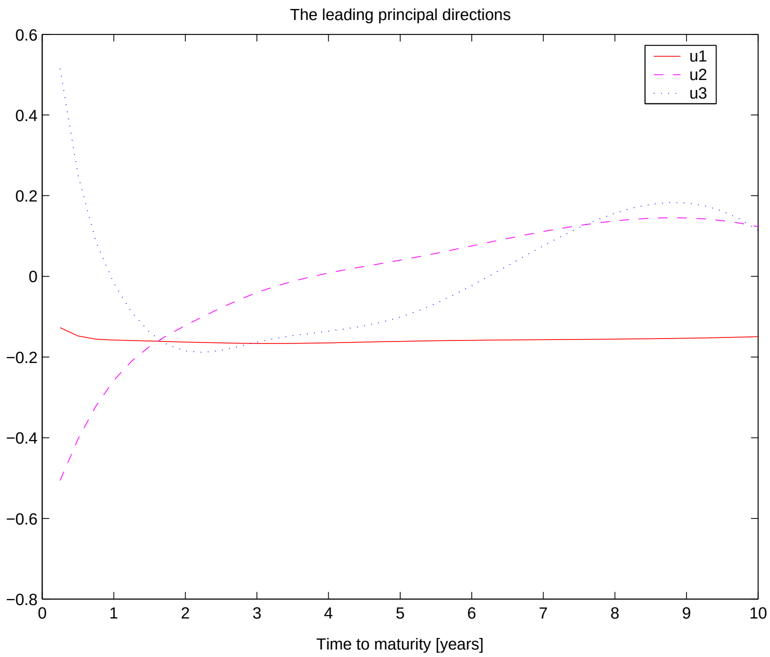



Figure 3: The first three principal components $u_1(\tau), u_2(\tau), u_3(\tau)$

These principal directions seem to agree with those found in earlier works (cf Litterman and Scheinkman (1991)). It is important to mention that the $u_i$ for $i = 1, 2, 3$ hardly depend on the time interval over which they are estimated. This is not the case for principal directions of higher order, and that is one good reason why one should avoid using models with more than three factors.

The first factor explains approximately 87% of the variance in the term structure. From the plot it is seen to be almost constant as a function of maturity, and hence it is responsible for parallel shifts of the yield curve. The second factor accounts for additional 10.5% of the variance and mainly tilts the curve. The third factor contributes additional 1.5%, and is responsible for changes in curvature.

Summarizing the above observations, the model should consist of at least two factors, but no more than three. A two-factor model produces higher excess yields, but at the expense of slightly higher variance of the residual noise. In this work we therefore focus on the three factor model, which makes the residual noise as small as possible without defining principal directions of higher order which are too sensitive to the selection of the learning period.

**Convenience yields in bond hedging.**

In the following examples we use the minimal number of bonds needed to hedge \$100 invested in a certain bond (which we index by $i_0$ in appendix C) against all factor risk. This is done by arbitrarily choosing different sets of bonds for the hedging. Since we use a three factor model ($F = 3$), we need *four* additional bonds to construct a hedged portfolio that costs zero. The bonds used for hedging were chosen arbitrarily, since the key issue here is not achieving big excess gains but rather demonstrating the good fit between theory and practice despite the existence of convenience yields that are substantially non-zero.

When analyzing empirical results we quantify the magnitude of excess gains by summing up monthly gains without discounting. In other words - we do not reinvest gains, 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 month from a zero net position. In the figures to follow, a dashed line stands for our theoretical prediction while solid lines show profit obtained in practice.

Figure 4 shows the cumulative profits from such portfolios, when hedging the 5-year note.

It is evident that the cumulative profit attained in practice (solid line) is in excellent agreement with what is predicted by our model (dashed line).

In addition, it is evident that the drift is substantially different from zero. For example, when using maturities of 3M,6M,3Y and 10Y, the null hypothesis is rejected with p-value of 0.05%, while the result is even more significant when hedging with 3M,6M,2Y and 3Y. Even in the case of hedging with 3M,6M,2Y and 10Y, where the profit plot seems close to zero, the p-value for rejecting the null hypothesis is about 10%.

While only \$100 were invested in the 5Y note, it turns out that in some cases bigger amounts are invested in other bonds that are used for the hedging. The hedging portfolios that correspond to Figure 4 are summarized in Table 1.

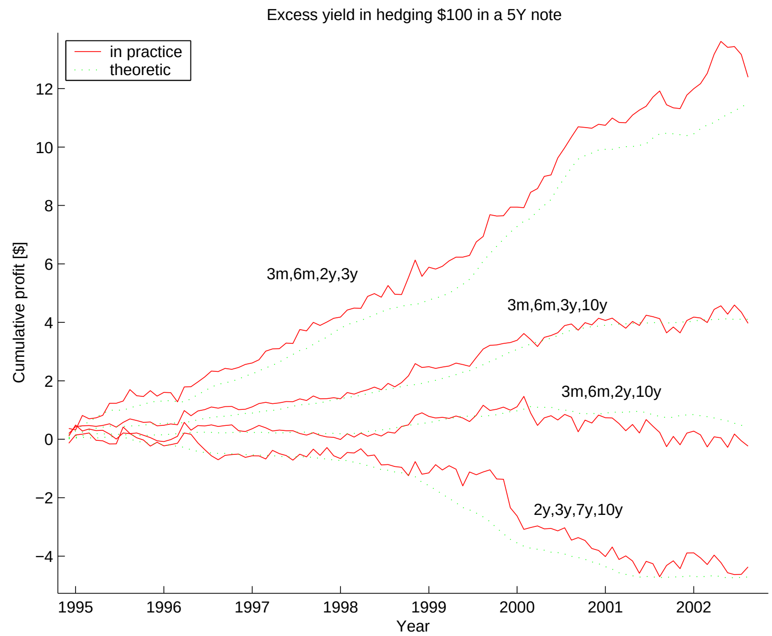



Figure 4: Hedging $100 invested in the 5Y note against three leading factors

| Maturity: | 3M | 6M | 1Y | 2Y | 3Y | 5Y | 7Y | 10Y |
|---|---|---|---|---|---|---|---|---|
| 3M,6M,2Y,3Y: | -129 | 66 | | 357 | -394 | 100 | | |
| 3M,6M,2Y,10Y: | -207 | 311 | | -178 | | 100 | | -26 |
| 3M,6M,3Y,10Y: | -181 | 229 | | | -131 | 100 | | -17 |
| 2Y,3Y,7Y,10Y: | | | | -17 | 4 | 100 | -127 | 40 |

Table 1: Description of the portfolios that hedge $100 in a 5Y note

A straightforward computation may provide a preliminary estimate of the magnitude of this phenomena. Consider, as an example, the first hedging portfolio that consists of maturities 3M, 6M, 2Y, and 3Y. In 7.5 years, the portfolio gained a profit of about $13, which is $1.73 per year on average. The net long and short positions in this portfolio are worth \$523 each, so the excess rate of return equals $1.73/523 \approx 0.33\%$ of the long (and short) size of the portfolio.

We hedged other bonds as well, always with similar results. In Figure 5 we provide one more result, this time for hedging of the 3M note using different sets of bonds.

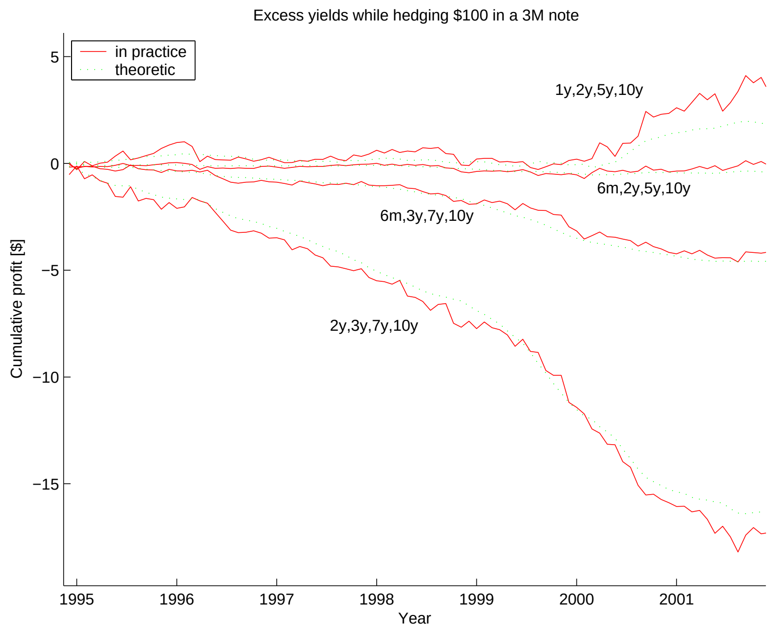



Figure 5: Hedging \$100 invested in the 3M note against three leading factors

Once again, it is clear that the choice of the hedging instruments may have a crucial effect on the excess return of the hedged portfolio.

The computation above serves as a preliminary estimate of the magnitude of this phenomena, with an arbitrarily chosen hedging portfolio, and with no attempt to optimize. Clearly optimization would increase these numbers substantially, as we now show.

**Maximizing the convenience yield.**

We now implement convenience yield optimization on the market data. In the figures below we plot cumulative profit from intertemporally maintaining an optimal portfolio. As before, we focus on out-of-sample testing, the plots sum up monthly gains without discounting, a dashed line stands for our theoretical prediction and a continuous line show profit obtained in practice.

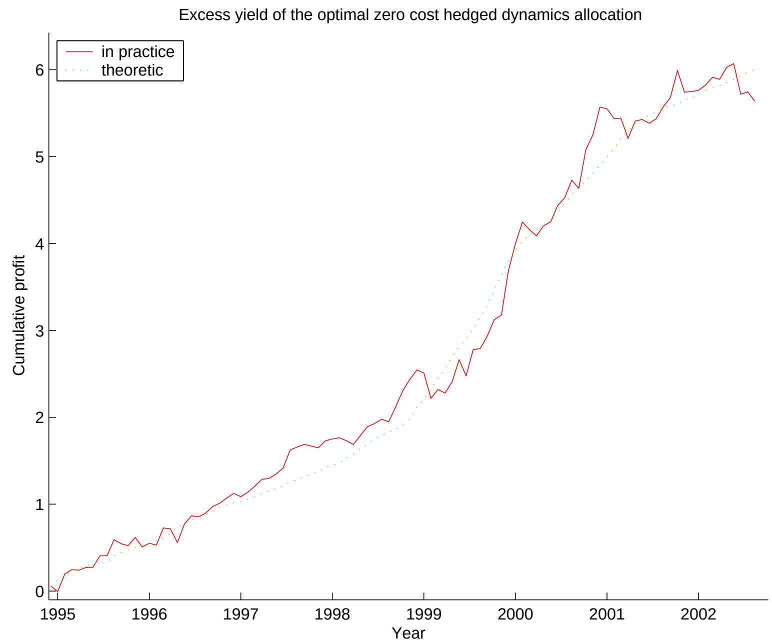



Figure 6: Arbitrage cumulative profit

Figure 6 shows the cumulative profit on a the optimal dynamic allocation that is kept zero cost with long size of \$100 and zero exposure to factor risk at all times. Once again, the excellent fit between the theory and practice is striking. In addition, a substantial gain is evident: we are able to generate zero cost dynamic allocation with low volatility return of about $6/7.5 = \$0.8$ per year, which is an excess yield of 0.8%.

It is important to mention that by considering only two factors rather than three, and furthermore by applying the theory to weekly rather than monthly data, this excess yield is more than doubled.

# 5 Convenience yield of individual bonds

So far we defined, quantified, and demonstrated the existence of convenience yields of portfolios with instantaneous zero variance. Here we ask what can be said about the excess yields of individual bonds.

This question is actually ill-posed, since the arbitrage is only relative: an arbitrage opportunity means that one bond is cheap relative to another, but it is impossible to determine which of these bonds are consistent with no-arbitrage, if any. In other words, there is no unique way to quantify the arbitrage in a specific bond. Nonetheless, in this section we suggest a definition of the convenience yield that best estimates the inconsistency of the bond yield with no-arbitrage, in the following sense:

It is well known (for example, Harrison and Kreps (1979)) that in a frictionless market there is no arbitrage iff there exists an equivalent measure in which all bond price processes become martingale. In a typical market with frictions however, this condition will not be satisfied, and there are non-zero convenience yields. Here we show that there is a unique way to define the convenience yield of specific bonds so that all convenience yield are zero iff there is no arbitrage (Theorem 5.1 below).

Denote by $\zeta_i(t)$ the orthogonal projection of the vector

$$\left(\frac{\mu(t,\tau_1)}{\tau_1}, \ldots, \frac{\mu(t,\tau_N)}{\tau_N}\right)$$

onto the subspace spanned by $\{(u_i(\tau_1), \ldots, u_i(\tau_N))\}_{i=1}^F$:

$$\zeta_i(t) = \left\langle \frac{\mu(t,\cdot)}{\cdot}, u_i(\cdot) \right\rangle \equiv \sum_{j=1}^N \frac{\mu(t,\tau_j)}{\tau_j} \cdot u_i(\tau_j), \qquad i = 1, \ldots, F. \tag{5.1}$$

We then denote by $\rho(t,\tau)$ the difference between $\mu(t,\tau)$ as defined in (2.2) and its projection:

$$\rho(t,\tau) := \mu(t,\tau) - \tau \sum_{i=1}^{F} u_i(\tau)\zeta_i(t), \tag{5.2}$$

and by $\rho(t)$ the vector of all $\rho_i(t) \equiv \rho(t,\tau_i)$ at a given $t$:

$$\rho(t) := (\rho_1(t), \ldots, \rho_N(t))^T.$$

Combining (2.1) with (5.2) we get:

$$\frac{dB(t,T)}{B(t,T)} - r(t)dt = \rho(t,\tau)dt + \tau \left[ \sum_{i=1}^{F} u_i(\tau)\left(\zeta_i(t)dt - dz_i(t)\right) - d\epsilon(t,\tau) \right]. \tag{5.3}$$

Eq. (5.3) and especially Theorem 5.1 below motivate the following definition:

**Definition.** The *convenience yield* at time $t$ of a bond maturing in $\tau$ years is $\rho(t,\tau)$ as defined in (5.2).

The following theorem (proven in appendix D) argues that there are no arbitrage opportunities if and only if all convenience yields are zero:

**Theorem 5.1** *Suppose that in (5.3) $d\epsilon(t,\tau) \equiv 0$, $\forall \tau$[3], $t$, namely,*

$$\frac{dB(t,T)}{B(t,T)} - r(t)dt = \rho(t,\tau)dt + \tau \left[ \sum_{i=1}^{F} u_i(\tau)\left(\zeta_i(t)dt - dz_i(t)\right) \right]. \tag{5.4}$$

[3]When a continuum rather than a finite set of maturities is considered, $dy$ may be decomposed into an infinite sum. Then the theorem still holds, but with $N$ replaced by $\infty$. In this case, the condition $\rho(t) = 0$ for $F < \infty$ must coincide with the Duffie and Kan (1996) condition. This condition can be derived by substituting $\mu$ from (2.2), $\zeta_i(t)$ from (5.1), and then $y$ from (2.3) into (5.2), while imposing $\rho(t,\tau) \equiv 0$. The significance of our approach is in realizing that typically this relationship does not hold in reality, thus leaving room for residual drifts which may be justified by market frictions.

*Then*

1. *If* $F < N$ *the model of (5.4) is arbitrage free iff* $\rho(t) = 0 \ \forall t$.

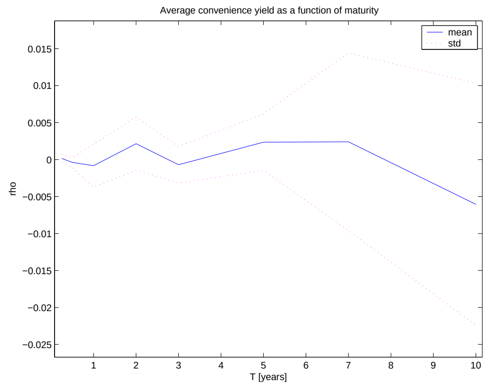
2. *If* $F = N$ *then* $\rho(t) = 0 \ \forall t$ *and the model of (5.4) is arbitrage free.*

The probability measure in which the processes $dz_i - \zeta_i dt$ become martingale is the closest thing to a risk neutral measure: it becomes the risk neutral measure when the market is arbitrage free.

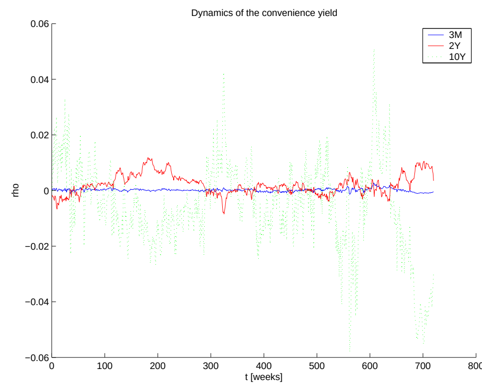
Theorem 5.1 asserts that $\rho_i(t)$ should vanish in a frictionless market. Hence, $\rho_i(t) < (>) 0$ implies a certain cost (reward) in holding security $i$. This motivates the terminology *convenience yield*: it is a certain cost (reward) for the relative convenience (inconvenience) associated with holding the security.

Note that our definition of the convenience yield is robust in the sense that it remains unchanged even if one fits a dynamic term structure model that predicts future yields. Any predictability in the leading factors of uncertainty contributes only to portfolios that contain these factors. By considering only portfolios in the subspace that is orthogonal to these factors, our convenience yield remains unaffected by such predictability. A more intuitive interpretation is that predictability of a certain bond is part of its compensation for risk, while $\rho(t,\tau)$ approximates the *risk-neutral* excess drift of the bond, which is the yield in excess of the compensation for risk.

Figure 7 summarizes properties of the convenience yields that we obtain for specific maturities. In Figure 7(a) average drifts as a function of maturity are introduced. The dotted lines denote a confidence interval of one twice the standard deviation for each maturity. In Figure 7(b) we plot the dynamic behavior of $\rho(t,\tau)$ for selected maturities. It is evident that the variance of the drift grows with maturity, and hence most of the time the long maturities produce larger convenience yields (in absolute value) than the short ones.




(a) Average $\rho$ and its standard deviation as a function of maturity




(b) $\rho$ as a function of $t$ for different maturities

Figure 7: Properties of $\rho$

# 6 Conclusions

We have proposed a new methodology in market modeling. In contrast with existing models, we do not assume that there are no arbitrage opportunities in the market, but rather acknowledge the possibility that market frictions may allow the existence of what would have been arbitrage opportunities if 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.

Our conclusion is that in order to fully characterize the price dynamics of any bond portfolio, in addition to its exposures to all factors of uncertainty as in Litterman and Scheinkman (1991), the convenience yield should be considered. In this work we provided the means to compute this quantity, and showed that the dynamics of bond prices indeed agrees with our theoretical observation.

A valid criticism would be that we trade bonds with fixed time to maturity. Such bonds typically do not exist, hence we trade interpolated prices. Indeed, one could improve the empirical study by trading only traded securities. Nonetheless, we have tested our method for robustness, so we expect to obtain essentially the same results.

In particular, we used different interpolation techniques for both determining bond prices and for differentiation of the yield curve. While some variability in the profit plots is evident, these changes are minor, and the excess yield remains significant. Clearly, since we construct zero cost portfolios, our results are completely insensitive to changes in the instantaneous rate $r(t)$.

In addition to this robustness which leads us to believe that our empirical results are trustworthy, the excellent fit between our model predictions and what is attained in practice remains as such in all robustness tests.

It would be interesting to explore the problem of pricing fixed income derivatives within this framework. Since non-zero convenience yields are present, the no-arbitrage assumption that lies in the heart of the pricing is violated by the underlying. However, in a market with $F$ factors of uncertainty there is an $F$-dimensional subspace of portfolios for which the convenience yield vanishes. This subspace may be considered as arbitrage free, and in this subspace there is an equivalent martingale measure, so risk neutral derivative pricing can be worked out.

In parallel research in progress, we have applied versions of our methodology to options markets and to futures markets with striking success. In other words, results of similar nature are common across all markets.

# 7 Appendix A - Principal Component Analysis (PCA)

We outline the general idea of PCA, with emphasis on the special features used in this paper. PCA dates back to Hotelling (1933). The interested reader may find more details in Jolliffe (1986).

Given a set of $N$-dimensional vectors $\{dy(t) \equiv (dy(t,\tau_1),\ldots,dy(t,\tau_N))^T\}$, $t = 1,\ldots,M$, PCA (sometimes referred to as the (discrete) Karhunen-Loeve transform) explains the maximum amount of variance possible by $F$ linearly transformed components. The $F$ *principal directions* $\{u_i\}$, $i = 1,\ldots,F$ are are those orthonormal axes onto which the data is projected.

Denote by $\overline{dy}$ the sample mean of $dy(t)$:

$$\overline{dy} := \frac{1}{M}\sum_{t=1}^{M} dy(t).$$

PCA can be defined recursively: define the direction $u_1$ of the first principal direction by:

$$u_1 := \arg\max_{\|u\|=1} \mathbb{E}[(u^T(dy(\cdot) - \overline{dy}))^2],$$

where $u$ is of the same dimension $N$ as the random data vector. (All the vectors here are column vectors). Thus, the first principal direction is the projection on the direction in which the variance of the projection is maximized. Having determined the first $k-1$ principal directions, the $k$-th principal direction is determined as the principal direction of the residual:

$$u_k := \arg\max_{\|u\|=1} \mathbb{E}\left[\left(u^T\left((dy(\cdot) - \overline{dy}) - \sum_{i=1}^{k-1} u_i u_i^T (dy(\cdot) - \overline{dy})\right)\right)^2\right].$$

In practice, the computation of the $u_i$ is accomplished using the (sample) covariance matrix

$$S := \mathbb{E}[(dy(\cdot) - \overline{dy})(dy(\cdot) - \overline{dy})^T] : \tag{7.5}$$

the $u_i$ are the eigenvectors of $S$ that correspond to its $F$ largest eigenvalues.

We denote $U := (u_1, \ldots, u_F)$ and $dz(t) := (dz_1(t), \ldots, dz_N(t))^T$. The factors (principal components) are then given by

$$dz(t) = U^T(dy(t) - \overline{dy}), \tag{7.6}$$

The basic goal in PCA is to reduce the dimension of the data. Thus one usually chooses $F << N$. Indeed, it can be proven [9] that the $F$ dimensional *principal subspace* spanned by $\{u_j\}$, $j = 1 \ldots, F$ is the one that minimizes the squared reconstruction error $\sum \|dy(t) - \tilde{dy}(t)\|^2$, where the optimal linear reconstruction of $dy(t)$ is given by

$$\tilde{dy}(t) = U dz(t) + \overline{dy}.$$

Such a reduction in dimension is important mainly since noise may be reduced, as the data not contained in the $F$ first components may be mostly due to noise.

Under ergodicity assumption, the factors are uncorrelated zero mean random variables. To see this, note that (7.6) implies:

$$\mathbb{E}[dz_i(\cdot)] = u_i^T \mathbb{E}[dy(t) - \overline{dy}] = 0,$$

and

$$\begin{aligned} \frac{1}{T}\sum_{i=1}^{T} dz(t)dz(t)^T &= \frac{1}{T}\sum_{i=1}^{T} U^T(dy(t) - \overline{dy})(dy(t) - \overline{dy})^T U \\ &= U^T \frac{1}{T}\sum_{i=1}^{T}(dy(t) - \overline{dy})(dy(t) - \overline{dy})^T U = U^T S U. \end{aligned} \tag{7.7}$$

Since $u_i$ are the eigenvectors of $S$, this product is a diagonal matrix, the elements of which are the $F$ largest eigenvalues, implying that the factors $dz_i$ are uncorrelated.

# 8 Appendix B - Proof of theorem 2.1

We use $T$ and $\tau$ interchangeably to denote maturity and time to maturity of the same bond, namely, $T \equiv t + \tau$. Denote by $\hat{y}(t,T)$ the yield to maturity (YTM) of the discount bond that matures at $T$. Unlike the definition $y(t,\tau)$ of the spot rate for $\tau$ years, here the time to maturity decreases as time passes. While $T \equiv t + \tau$ implies $\hat{y}(t,T) = y(t,\tau)$, the dynamics of these quantities differ: $d\hat{y}(t,T) \neq dy(t,\tau)$. The following lemma quantifies this difference:

**Lemma 8.1**

$$d\hat{y}(t,T) = dy(t,\tau) - \frac{\partial y(t,\tau)}{\partial \tau} dt, \tag{8.8}$$

*where $\partial y(t,\tau)/\partial\tau$ is the slope of the term structure at $\tau$ when $t$ is fixed, and $dy(t,\tau)$ is specified in (2.3).*

**Proof.** Unlike the case when $\tau$ is constant, when $T$ is fixed $dy(t,\tau)$ depends on $t$ also via $\tau$. Hence,

$$d\hat{y}(t,T) \equiv dy(t,T-t) = \frac{\partial y(t,\tau)}{\partial t} dt + \frac{\partial y(t,\tau)}{\partial \tau} d\tau = dy(t,\tau) - \frac{\partial y(t,\tau)}{\partial \tau} dt. \tag{8.9}$$

□

Using Ito's Lemma,

$$\begin{aligned}
\frac{dB(t,T)}{B(t,T)} &- r(t)dt = -d\left[(T-t)\hat{y}(t,T)\right] + \frac{1}{2}(T-t)^2 d\langle \hat{y}(t,T)\rangle - r(t)dt \\
&= [\hat{y}(t,T) - r(t)]dt - (T-t)d\hat{y}(t,T) + \frac{1}{2}(T-t)^2 d\langle \hat{y}(t,T)\rangle \\
&\equiv \mu(t,T-t)dt - (T-t)\left[\sum_{i=1}^{F} u_i(T-t)dz_i(t) + d\epsilon(t,T-t)\right], \quad (8.10)
\end{aligned}$$

with $\mu(t,\tau)$ as in eq. (2.2).

# 9 Appendix C: Zero cost risk free portfolios

## 9.1 Static risk free portfolios

We use the minimal possible number of bonds to generate a zero cost portfolio with no factor risk, by satisfying eq. (9.11) for all $k \in \{1, \ldots, F\}$:

$$\sum_{i=1}^{N} \pi_i(t)\tau_i u_k(\tau_i) = 0. \tag{9.11}$$

Note that since this set of equations is homogeneous, if there are $F$ factors of uncertainty, at least $F+1$ assets are needed to eliminate all factor risk, and then the risk free portfolio is determined up to a multiplicative constant. In addition we would like to avoid using the risk free rate when evaluating portfolio performance, and this is achieved by using an additional bond and adding the constraint that the portfolio costs zero:

$$\sum_{i=1}^{N} \pi_i = 0.$$

Having selected $(F+2)$ bonds, indexed by $i_0, \ldots, i_{F+1}$, we choose to invest $\pi_0 = \$1$ in bond $i_0$ in order to remain with a unique solution. Then we have $(F+1)$ linear equations with the same number of unknowns: $\pi := (\pi_{i_1}, \ldots, \pi_{i_{F+1}})^T$.

Denote

$$W := \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \tau_{i_1}u_1(\tau_{i_1}) & \tau_{i_2}u_1(\tau_{i_2}) & \ldots & \tau_{i_{F+1}}u_1(\tau_{i_{F+1}}) \\ \vdots & & & \\ \tau_{i_1}u_F(\tau_{i_1}) & \tau_{i_2}u_F(\tau_{i_2}) & \ldots & \tau_{i_{F+1}}u_F(\tau_{i_{F+1}}) \end{pmatrix}, \tag{9.12}$$

and

$$v := -\begin{pmatrix} 1 \\ \tau_{i_0} u_1(\tau_{i_0}) \\ \vdots \\ \tau_{i_0} u_F(\tau_{i_0}) \end{pmatrix}. \tag{9.13}$$

Then $\pi = W^{-1}v$ is a zero cost portfolio with no factor risk, and its convenience yield is $\pi^T \rho$.

## 9.2 An optimal portfolio

We would like to solve an optimization problem of the form: assuming (5.4) holds, for an investor with no wealth at time $t$, find a normalized portfolio $\pi(t)$ with no factor risk for which the expected profit rate is maximized. Formally,

$$\max_{\pi(t)} \sum_{i=1}^{N} \pi_i(t)\rho_i(t) \tag{9.14}$$

s.t.

$$\begin{array}{lll} \sum_{i=1}^{N} \pi_i(t) = 0 & & (\textit{zero net investment}) \\ \sum_{i=1}^{N} \pi_i(t)\tau_i u_j(\tau_i) = 0, & j = 1, \ldots, F & (\textit{zero factor risk}) \\ \|\pi(t)\| = 1 & & (\textit{normalization}). \end{array} \tag{9.15}$$

The norm condition is aimed at limiting the size of the long and short sides of the portfolio, and should be linked to margin requirements, but then the solution must be numeric. In the case of Euclidean norm, $\|\pi(t)\| := \sqrt{\sum_{i=1}^{N} \pi_i^2(t)}$, this problem can be solved analytically, so this is what we assume in order to proceed. Once the Lagrangian is introduced, the computation is straightforward and therefore omitted:

**Proposition 9.1** *The optimal portfolio for the optimization problem (9.14), (9.15) is:*

$$\pi_{opt}(t) = \frac{\left[I - W^T(WW^T)^{-1}W\right]\rho(t)}{\|\left[I - W^T(WW^T)^{-1}W\right]\rho(t)\|}, \tag{9.16}$$

*where*

$$\rho(t) := (\rho_1(t), \ldots, \rho_N(t))^T,$$

$$W := \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \tau_1 u_1(\tau_1) & \tau_2 u_1(\tau_2) & \ldots & \tau_N u_1(\tau_N) \\ \vdots & & & \\ \tau_1 u_F(\tau_1) & \tau_2 u_F(\tau_2) & \ldots & \tau_N u_F(\tau_N) \end{pmatrix}, \tag{9.17}$$

*and $I$ is the identity matrix.*

# 10 Appendix D: Proof of Theorem 5.1

It is convenient to denote $w_j := (\tau_1 u_j(\tau_1), \ldots, \tau_N u_j(\tau_N))$, $j = 1, \ldots, F$.

If $F < N$, unless $\mu(t,\tau)/\tau$ is in the subspace spanned by $\{u_i(\tau)\}_{i=1}^F$, $\rho(t) \neq 0$ and it is orthogonal to the subspace spanned by $\{w_j\}$, $j = 1, \ldots, F$.

Consider a portfolio $\pi(t) \equiv (\pi_1(t), \ldots, \pi_N(t))^T$ at time $t$ (recall that $\pi_i(t)$ ($i = 1, \ldots, N$) is the amount of money held at time $t$ in bond of maturity $T_i$), which satisfies

$$\langle \pi(t), w_j \rangle = 0, \quad j = 1, \ldots, F. \tag{10.18}$$

Then the portfolio $\pi(t)$ is risk free, and the collection of all $\pi(t)$ satisfying (10.18) form the complementary subspace orthogonal to the subspace spanned by $\{w_j\}_{j=1}^F$. Thus, one may choose $\pi(t)$ such that $\langle \pi(t), \rho(t) \rangle \neq 0$.

Then, (5.4) implies that this portfolio is risk free, and yet yields a rate of return that exceeds the risk free rate, $r(t)$, and hence the existence of arbitrage.

For the case $F = N$, the orthonormal basis $u_i(\tau)_{i=1}^N$ spans the whole space and therefore $\rho(t) \equiv 0 \; \forall t$. A risk free portfolio must satisfy (10.18), and then (5.4) implies it yield zero return and hence no arbitrage.

□

# References


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

[2] Diebold, F.X. and C. Li (2002) “Forecasting the term structure of government bond yields”, *PIER Working Paper 02-026.*

[3] Dai, Q. and K. Singleton (2003), “Expectation puzzles, time-varying risk premia, and dynamic models of the term structure”, forthcoming *Journal of Financial Economics.*

[4] Fama, E. (1976), “Forward rates as predictors of future spot rates”, *Journal of Financial Economics 3, 361-377.*

[5] Heath D., R. Jarrow and A. Morton (1992) “Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuation ”, *Econometrica 60, 1*, 77-105.

[6] Harrison, J.M. and D.M. Kreps, (1979), “Martingales and arbitrage in mutliperiod securities markets”, *Journal of Economic Theory 20*, 381-408.

[7] Hotelling, H. (1933), “Analysis of a complex of statistical variables into principal components”, *Journal of Educational Psychology 24*, 417-441.

[8] Hull, J.C. (1997), “Options, Futures, and other Derivatives”, Third ed., *Prentice-Hall.*

[9] Jolliffe, I.T. (1986) *Principal Component Analysis* Springer-Verlag.

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

[11] Longstaff, F.A. and E.S. Schwarts (1992), “Interest-Rate volatility and the term structure: a two-factor general equilibrium model”, *The Journal of Finance, 47(4)*, 1259-1282.

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