Universität St.Gallen

Statistical arbitrage in the U.S. treasury futures market

Wale Dare

September 2017 Discussion Paper no. 2017-16

School of Economics and Political Science,
Department of Economics

University of St.Gallen

Editor:	Martina Flockerzi University of St.Gallen School of Economics and Political Science Department of Economics Müller-Friedberg-Strasse 6/8 CH-9000 St. Gallen Phone +41 71 224 23 25 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Müller-Friedberg-Strasse 6/8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	http://www.seps.unisg.ch

Statistical arbitrage in the U.S. treasury futures market

Wale Dare

Author's address: Wale Dare
Faculty of Mathematics and Statistics
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 224 2413
Email wale.dare@unisg.ch

Abstract

We argue empirically that the U.S. treasury futures market is informational inefficient. We show that an intraday strategy based on the assumption of cointegrated treasury futures prices earns statistically significant excess return over the equally weighted portfolio of treasury futures. We also provide empirical backing for the claim that the same strategy, financed by taking a short position in the 2-Year treasury futures contract, gives rise to a statistical arbitrage.

Keywords

Market efficiency, U.S. treasury futures, statistical arbitrage, joint-hypothesis.

JEL Classification

C12, G13, G14.

1 Introduction

Is the U.S. treasury bond futures market informational efficient? Weak-form informational efficiency requires all strategies that rely solely on historical price data to be dominated by the passive strategy of holding single traded assets or a weighted portfolio of traded assets. The notion of dominance as it relates to asset pricing was introduced by Merton (1973) to study option pricing formulas that are consistent with rational investor behavior. More recently, Jarrow & Larsson (2012) obtained a characterization of informational efficiency in terms of the no dominance condition (ND) and the No Free Lunch with Vanishing Risk condition (NFLVR) of Delbaen & Schachermayer (1994). Accordingly, market *inefficiency* can be asserted as soon as either the ND or NFLVR fails.

This result simplifies considerably the task of verifying market efficiency; it belies the long held belief that in order to test for violations of market efficiency, one must first specify a model of equilibrium prices such as the CAPM and then test for efficiency in relation to the estimated equilibrium model. Unfortunately, this two step procedure runs quickly into difficulties, since it may not be possible to tell apart errors due to model misspecification and those that are solely due to market inefficiency. This is the well-known joint-hypothesis problem discussed in (Fama, 1969).

Moreover, the No Dominance condition itself could be dispensed with as soon as a change of numeraire is performed. Indeed let $\mathcal{B} := (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ denote a probability basis, and let S denote an n-dimensional semimartingale whose components $S^i, 0 \leq i < n$, represent the price of n distinct assets, expressed in units of the zeroth asset. For the sake of convenience, also assume that at time zero, each asset is priced at one, i.e. $S^i_0 = 1$ for $0 \leq i < n$. Now, let γ denote a positive number between zero and one, i.e. $0 < \gamma < 1$, and define

$$Z^{\gamma,i} := (S, S^{\gamma,i})(S^{\gamma,i})^{-1},$$

where $S^{\gamma,i} = \gamma + (1-\gamma)S^i$. According to Dare (2017, Proposition 2.1), the efficiency of $(S, \mathcal{B})$ is equivalent to the existence of a local martingale measure for the markets $(Z^{\gamma,i}, \mathcal{B})$, for $0 \leq i < n$ and $0 < \gamma < 1$.

In fact, a stronger statement can be made achieve a slightly provided prices are expressed in units of a portfolio constructed on the basis of a strictly positive weight vector $\alpha = (\alpha_0, \cdots, \alpha_{n-1})$, i.e. $\alpha_i > 0$ for $0 \leq i < n$. Indeed if

$$Z^{\alpha} := (S, S^{\alpha})(S^{\alpha})^{-1},$$

then according to Dare (2017, Corollary 2.2), the market $(S, \mathcal{B})$ is efficient if and only if $(Z^\alpha, \mathcal{B})$ admits a local martingale measure. The choice of a market portfolio is irrelevant so long as it assigns positive weight to each traded asset.

We will argue for a violation of market efficiency using Dare (2017, Proposition 2.1) with S^i representing the price of the 2-Year U.S. Treasury futures contract. Fortunately, since the NFLVR condition is specified in terms of the physical measure, the joint-hypothesis issue may be avoided by evaluating trading rule for violations of NFLVR. Using this testing approach, we make and emprically support the claim that between April 1, 2010 and December 31, 2015, the equally weighted buy-and-hold strategy was out-performed by a simple cointegration-based trading rule. Moreover, the hypothesis of the existence of a statistical arbitrage, in the sense of (Hogan et al., 2004), achieves a p-value less than 2%.

The trading rule we examine takes as starting point the hypothesis that treasury bond futures are cointegrated and then attempts to profit from deviations from the cointegrating relationships. The cointegration hypothesis assumes, among other things, that even though prices of individual contracts may be non-stationary, there exists at least one linear combination of these contracts that results in a stationary price process. That is to say, it is possible to put together a portfolio of long and short positions in individual contracts such that the resulting market value of the portfolio is stationary. The hypothesis of cointegrated bond prices has been examined by Bradley & Lumpkin (1992), Zhang (1993), and many others. In these studies, the data employed was sampled at low frequency, daily or monthly, and the hypothesis of integrated bond prices could not be rejected . We carry out similar analysis and find empirical support for cointegration using data sampled intra day at one-minute intervals.

We obtain theoretical motivation for the cointegration-based trading rule by embedding our analysis within the literature devoted to the study of the term structure of bonds using factor models. Starting with Litterman & Scheinkman (1991) and later Bouchaud et al. (1999) and many others, it has been noted that between 96% and 98% of overall variance of the entire family of treasury securities may be explained by the variance of just three factors, the so-called level, slope, and curvature factors. The factors are so named because of how they affect the shape of the yield curve. A shock emanating from the first factor has nearly the same impact on contracts of all maturities; the resulting effect is a vertical shift, upward or downward, of the entire yield curve. The second factor affects bonds of different maturities in such a manner as to change the steepness or slope of the curve; it does so by affecting securities at one end of the maturity spectrum more or less than those at the other end. Finally, the third factor has the effect of making the yield curve curvier; it does so by having more or less pronounced effects on medium term bonds than on bonds situated either ends of the maturity spectrum.

We argue that a strategy based on a cointegration hypothesis is natural within the context of a term structure driven by common stochastic trends or factors. In fact, the opposite is also true, that is, a common factor structure is a natural consequence of cointegrated yields. This line of argument provides support based on economic theory for our strategy and helps explain its performance. Our results suggests that the futures market may be inefficient. Market inefficiency is clearly not a desired outcome. It implies the existence of a free lunch. Put another way, our results points to possible misallocation of resources.

The rest of the paper proceeds as follows: in section 2, we provide a description of the data used. Futures price data usually does not come in continuous form for extended periods of time, so we had to make certain choices about how available historical price data is transformed into a state suitable for our analysis. These choices can be implemented in real-time and are, therefore, to be considered as part of the trading rule. In section 3, we provide theoretical foundation for our trading rule. This foundation allows us to reach beyond our data and assert that the profitability of the trading rule is very likely not confined to the period for which we have data. Section 4 is devoted to the implementation details of the trading rule. Section 5 summarizes our empirical results, and section 6 concludes.

2 Data

2.1 Treasury futures

CBOT Treasury futures are standardized foreward contracts for selling and buying US government debt obligations for future delivery or settlement. They were introduced in the nineteen-seventies at the Chicago Board of Trade (CBOT), now part of the Chincago Merchantile Exchange (CME), for hedging short-term risks on U.S. treasury yields. They come in four tenors or maturities: 2, 5, 10, and 30 years. In reality, each contract type is written on a basket of U.S. treasury notes and bonds with a range of maturities and coupon rates. For instance, the 30-Year Treasury Bond Futures contract is written on a basket of bonds with maturities ranging from 15 to 25 years. It is, therefore, worth keeping in mind that a study of the dynamics of the yield curve using futures data reflects influences from a range of maturities.

Every contract listed above except the 2-Year T-Note Futures contract, which has a face value of \$200,000, has a face value of \$100,000. That is each contract affords the buyer the right to buy an underlying treasury note or bond with a face value of \$100,000 or \$200,000 in the case of the 2-Year contract. In practice, the price of these contracts are quoted as percentages of their par value. The minimum tick size of the 2-Year T-Note Futures is 1/128%, that of the 5-Year T-Note Futures is 1/128%, that of the 10-Year T-Note Futures is 1/64%, and that of the 30-Year T-Bond Futures contract is 1/32%. In Dollar terms, this comes to \$15.625, \$7.8125, \$15.625, and \$31.25, respectively, per tick movement. [1] These tick sizes are orders of maginitude larger than those typically encounted in the equity markets.

Even though most futures contracts are settled in cash at the expiration of the contract, for a small percentage of open interests, delivery of the underlying bond actually takes place. Given that the futures contract is written on a basket of notes and bonds, the actual bond or note delivered is at the discretion of the seller of the contract. In practice, the seller merely selects the cheapest bond in the basket to deliever. For our purposes, we shall focus on only the above listed tenors, but it is worth keeping in mind that there is is also a 30-Year Ultra contract that is also traded at the CME.

For our analysis, we use quote data, prices and sizes, from April 1, 2010 through December 31, 2015. Even though we have at our disposal data rich enough to allow resolution down to the nearest millisecond, we opted, arbitrarily, to aggregate the data into one-minute time bars. The representative quoted price and size for each time bar is the last recorded quote falling within that interval. Our use of quotes , bids and offers, instead of transaction data allows the computation of a proxy for the unobserved *true* price, by means of the mid-quote, at a higher frequency than transaction prices might have allowed. Using quotes, we are also able to reflect directly a major portion of the execution costs associated with any transaction, i.e. the bid-ask spread.

Trading in these markets primarily takes place electronically via CME ClearPort Clearing virtually around the clock between the hours of 18:00 and 17:00 (Chicago Time), Sunday through Friday. But, the markets are at their most active during the daytime trading hours of 7:20 and 14:00 (Chicago Time), Monday through Friday. This also the opening hours of the open outcry trading pits. For our analysis, We use exclusively data from the daytime trading hours. This ensures that the strategy is able to benefit from the best liquidity these markets can offer, while mitigating the effects of slippage (orders not getting filled at the stated price) and costs associated

[1] We refer the reader to more detailed information about the features of each contract to Labuszewski et al. (2014).

with breaking through the Level 1 bid and ask sizes.

2.2 Continuous prices

Unlike stocks and long bonds, futures contracts tend to be short-lived, with price histories extending over a few weeks or months. This stems from the traditional use of futures contracts as short-term hedging instruments against price/interest rate fluctuations. Treasury futures contracts, in particular, have a quarterly expiration cycle in March, June, September, and December. At any given point in time, several contracts written on the same underlying bond, differentiated only by their expiration dates, may trade side by side.

Usually, the next contract due to expire, the so-called front-month contract, offers the most liquidity. As the front-month approaches expiration, liquidity is gradually transferred to the next contract in line to expire, the deferred month contract. At any rate, a given contract is only actively traded for a few months or weeks before it expires. Hence, holding a long-term position in a futures contract actually entails actively trading in and out of the front month contract as it nears its expiration date. The implementation of this process is known as rolling the front month forward.

For the purpose of evaluating a trading strategy over a historical period of more than a few months, the roll can be retroactively implemented to generate a continuous price data. The usual way to go about the roll is to trade out of the front-month a given number of days before it expires. In the extreme case, the roll takes place on the expiration date of the front month contract. The downside of this type of approach is that the roll may take place at a date when liquidity in the deferred month is not yet plentiful. The result is that a backtest may not necessarily capture the increased trading cost associated with the lower liquidity level.

Our preferred approach for implementing the roll is to start trading out of the front month contract at any point during its *expiration month* as soon as the open interest in the deferred month contract exceeds the open interest in the front month contract. The data used in our backtest is spliced together this way; the procedure is implementable in real-time and must be considered part of the trading strategy discussed in this paper.

Now, while retroactive contract rolling may solve the problem of creating an unbroken long-term price history, it creates another: splicing prices together as described above would invariably introduce artificial price jumps into historical prices. To see this, consider a futures contract with price F written on a bond with price B. Using an arbitrage argument and ignoring accrued interest, the price of a futures contract at any time t may be expressed as:

$$F_t = B_t e^{(r-c)d}, \tag{2.1}$$

where c is the continuously compounded rate of discounted coupon payments on the underlying bond, d is the number of time units before the futures contract expires, and r is the repo rate.

Now, assuming the roll takes place in the expiration month, d for the front month is less than 30 days, whereas for the deferred month contract, d is at least 90 days. This results in a price differential between the two contracts, which shows up in the price data as a jump. In reality, and assuming a self-financing strategy, the price differential would necessitate a change in the number of contracts held, so that overall, the return on the portfolio is unaffected by the roll. Hence, in order to avoid fictitious gains and losses, the price series must be adjusted to remove the roll-induced price jumps.

The most often used methods in practice applies an adjustment to prices either prior or subsequent to the roll date. When the adjustment is applied to prices recorded after the contract is rolled forward, the price history is said to be adjusted forward; if on the other hand, the adjustment is applied to prices recorded prior to the roll date then the prices are said to be adjusted backward. The actual price adjustment, in the case of a backward adjustment, is most commonly carried out in one of two ways: in the first instance, the roll-induced price gap (price after roll minus price right after roll) is subtracted from all prices recorded prior to the roll date; in the second instance, all prices preceding the roll date are multiplied by a factor representing the relative price level before and after the roll. The second approach is reminiscent of how stock prices are adjusted after a stock split. We will refer to the first approach as the *backward difference adjustment method* and to the second as the *backward ratio adjustment method.* Forward ratio adjustment and forward difference adjustment are implemented similarly with the adjustments applied to prices recorded after the roll date. In our analysis, we will only consider backward adjusted prices, as they appear to be the more intuitive approach.

Both types of backward price adjustment methods are widely-used in practice, but the ratio adjustment method has the advantage of guaranteeing that prices, however early in the price series, always remain positive. In theory, the difference adjustment approach may generate negative prices given enough roll-induced price gaps. We mention these adjustment procedures because they tend to affect the performance of most strategies, including the one we study in this paper. The price adjustment procedures cannot be considered as part of a real-time trading strategy, so we report results using both the backward ratio adjustment and the backward difference adjustment.

3 Economic framework

3.1 The price of a futures contract

The traditional way of pricing a futures contract is via an arbitrage argument. The argument is best illustrated via an example. Suppose an agent, at time t (today), has a need to purchase a 10-Year Treasury note at time T_1. That is, at time T_1 when the forward/futures contract expires, the treasury note will mature in ten years at time $T_1 + 10$. The agent could go about it by borrowing money at time t at the repo rate to cover the full price of the bond. The full price of the bond would include the current spot price of the bond and accrued interest on the bond since the last coupon payment. The accrued interest is the portion of the next coupon payment that is due to the previous owner of the bond. Let's denote the spot price of the bond by B_t and the accrued interest by I_t. So, at time t the agent may borrow $B_t + I_t$, using the bond as collateral against the loan.

At time T_1, the loan used by the agent to fund the purchase would have accrued interest of its own and would have grown to $(B_t + I_t)e^{r(T_1-t)}$. Here, we are assuming a fixed repo rate r. On the other hand, taking procession of the treasury note endows the agent with the right to receive coupon payments generated by the note. Coupon rates are usually a fixed percentage of the par value of the bond. In practice, this is usually around 6% and payable semiannually; for this illustration, we will imaging that the coupon payments are payed continuously at the instantaneous rate of c. To recap, at time T_1, the loan balance grows to $(B_t+I_t)e^{r(T_1-t)}$, but it is offset by coupon payments of $Ke^{-c(T_1-t)}$, where K denotes the par value of the bond. Hence, at time T_1, for the agent to own the treasury bond outright, she simply needs to repay the loan, but because of the accrued interets she would only be out of pocket $F := (B_t + I_t)e^{r(T_1-t)} - Ke^{-c(T_1-t)}$. Hence, at time t it only makes economic sense to enter a futures contract if it is priced in such a way as to equate the cost of replicating it, that is, F.

The above analysis demonstrates that the price of a futures contract may be written in terms of the price of the underlying bond. In fact, by denoting C the continuously discounted present value of all coupon payments generated by the bond, we may write:

$$F_t = (B_t - C)e^{rd}, \tag{3.2}$$

where d is the amount of time left before the futures contract expires. It is worth noting that the foregoing analysis relies on the assumption of constant interest rate. It is also to be noted that the price of a futures contracts using a no-arbitrage argument may differ from the price of a forward contract in an environment with stochastic time varying interest rates. We refer the reader to Cox et al. (1981) for a lucid discussion of this point. For the intuition we wish to develop, the assumption of a constant in time interest rate is tolerable.

Returning to (3.2), and taking the natural logarithm of both sides of the equation, and assuming that the face value of the bond is much larger than the present value of future coupon payments, we may write

$$f_t \approx b_t - c + rd, \tag{3.3}$$

where $b_t := \log(B_t)$ and $c := \log(C)$. Note that $b_t = -Ty_t$, where y_t is the yield to maturity at time t of the bond. The quantity $rd - c$ is usually referred to variously in the empirical literature as the *carry* or the *basis*. Looking at actual price data, the carry would fluctuate from time to time usually around a long term mean. The basic idea of a mean-reverting carry is the motivation behind the so-called carry-trade, which is implemented by going short the futures and long the bond when the carry is high and doing the opposite when the carry is deemed too low. The strategy reviewed subsequently, is only related to this trade by the fact that it relies on mean-reversion to be profitable.

Returning to (3.3) it is apparent that besides the variation in the carry, variations in the logarithm of the futures price comes about because of variations in the logarithm of the bond price, which is itself driven by the yield to maturity of the bond. Usually, the carry does not vary by a whole lot and it is often modeled as a constant as we have done here unless, of course, the object of the analysis is to study the carry itself. Given these considerations we may model the logarithm of the price of the futures contract directly and exclusively in terms of the yield to maturity with no significant loss in rigor. That is, we may write

$$f_t = \alpha + \beta y_t, \tag{3.4}$$

where α and β are constant terms and y_t is the yield to maturity of the underlying bond. The constant α is simply the carry and whatever needs to be added or subtracted in order to make the approximation in (3.3) an equality. The constant β is in this setting equal to $-T$, that is, negative the tenor of the underlying bond.

The preceding reformulation of the logarithm of the price of a futures contract in terms of the yield to maturity of the underlying bond allows us to use the theoretical machinery developed to study the term structure of interest rates to motivate the trading system that we discuss subsequently.

3.2 Factor model of the yield curve

Factor modeling of the yield curve has a rich history in the financial literature. The extant models may be broadly classified under three main headings: statistical, no-arbitrage, and hybrid models. The static Nelson & Siegel (1987) (NS) model of the yield curve and its modern counterpart, the Dynamic Nelson-Siegel (DNS) model, proposed by Diebold & Li (2006) are prototypes of the class of statistical factor models of the interest rate term structure. They, especially the static Nelson & Siegel, are widely used both by financial market practitioners and central banks to set interest rates and forecast yields. Despite their popularity and appealing statistical properties, they tend to give rise to violations of the no-arbitrage condition[2].

The dynamic term structure models (DTSM) studied in (Singleton, 2006, Chapter12), of which the yield-factor model of Duffie & Kan (1996) is an early example, constitute the class of arbitrage-free models. These models derive a functional form of the yield curve in terms of state variables or factors, which also govern the market price of risk linking the local martingale measure to the historical measure. They are, therefore, by construction arbitrage-free. Despite their economic soundness, these models tend to have sub-par empirical performance. For instance, Dybvig et al. (1996) showed in the discrete-time setting that in an arbitrage-free model, long forward and zero-coupon rates can never fall; working in the general setting of continuous trading, Hubalek et al. (2002) arrived at a similar conclusion regarding the monotonicity of long forward rates under the no-arbitrage assumption. Clearly, this implication of the no-arbitrage framework is often contradicted by the empirical evidence that zero-coupon rates do in fact fall. Furthermore, negative rates and unit roots are ruled out. As (Diebold & Rudebusch, 2013, p. 13) put it

> Economic [no-arbitrage] theory strongly suggests that nominal bond yields should not have unit roots, because the yields are bounded below by zero, whereas unit root processes have random walk components and therefore will eventually cross zero almost surely.

Negative interest rates post 2008 financial crisis are a mainstay of many developed economies, including Switzerland. Moreover, the task of fitting arbitrage-free models to interest rate data can be very difficult since they tend to be over-parametrized and, typically, would generate multiple likelihood maxima (Diebold & Rudebusch, 2013, p. 55).

[2] See (Filipović, 1999) for such violations in the case of DNS models.

Lastly, the Arbitrage-Free Nelson-Siegel (AFNS) model proposed by Christensen et al. (2011) is a prototype of the hybrid class of models. It maintains the parsimonious parametrization of the DNS model while remaining arbitrage-free. The AFNS differs, at least in the functional form of the yield curve, from the standard DNS model only by the inclusion of an extra term known as the "yield adjustment factor". Intuitively, the AFNS model may be thought of as the projection of an arbitrage-free affine term structure model, namely the Duffie & Kan (1996) model, onto the DNS model with the orthogonal component swept into the yield adjustment factor.

The factor models briefly surveyed above motivate the trading rule adopted in this paper; it relies on the hypothesis that the term structure of interest rates can be described by an affine function of a set of state variables, notably the level, slope, and curvature principal components. Moreover, there is ample empirical evidence suggesting that the term structure is cointegrated. In particular, using monthly Treasury bill data from January 1970 until December 1988, Hall et al. (1992) observed that yields to maturity of Treasury bills are cointegrated and that during periods when the Federal Reserve specifically targeted short-term interest rates, the spreads between yields of different maturities defined the cointegrating vector.

In general, given a $N \in \mathbb{N}$ bonds, with N not necessarily finite, a factors model of the yield curve would represent the yield on the i-th bond as:

$$y_{i,t} = \alpha_t + \sum_{j=1}^{q} \beta_{i,j} f_{j,t} + \varepsilon_{i,t}, \tag{3.5}$$

where α is deterministic, q is a small number, f_j for $j = 1, \cdots, q$, are factors $\beta_{i,j}$ is the contribution of the jth factor to the ith bond, and ε_i is the component of the ith bond that is apart from any other bond. For our purposes, it does not actually matter whether the factors are macroeconomic or statistical in nature, but to fix ideas we assume $q = 3$ and the factors are the *level, slope*, and *curvature* factors of Litterman & Scheinkman (1991). By substituting the expression in (3.5) into equation (3.4), we obtain the log futures price in terms of the level, slope, and curvature of the term structure. That is,

$$f_{i,t} = \mu_t + \sum_{j=1}^{3} \gamma_{i,j} f_{j,t} + \varepsilon_{i,t}.$$

3.3 Factor extraction

Using Principal Component Analysis (PCA), it is possible to transform the original time series of futures prices into a set of orthogonal time series known as principal components. Because of the orthogonality property, the original time series may be expressed uniquely as a linear combination of the principal components. This representation motivates the interpretation of the principal components as the latent risk factors driving observed price fluctuations.

The analysis starts with n observations from an m-dimensional random vector, the original time series data. Then, assuming that the original time series admits a stationary distribution with finite first and second moments, the covariance matrix is estimated using an unbiased and consistent estimator. In our setting, the assumption of stationarity applied directly to the logarithm of futures prices is hard to justify. Prices generally trend upward, and the same may be expected for their log-transformed versions. Using the Augmented Dickey-Fuller (ADF) statistics with constant drift, we test the hypothesis that the lag polynomial characterizing the underlying data generating process has a unit root.

A quick scan of Table 1 reveals that for the most part the unit root assumption cannot be rejected. The only exception seems to be the 2 Year and the 5 Year futures price data for the year 2015, for which the assumption of a unit root may be rejected at the 5% confidence level. We think this outcome is a temporary fluke since for the previous five years the null hypothesis could not be rejected. We have also looked at different subsamples of the 2015 data, and for the most part the assumption of a unit root could not be rejected.

Under the circumstances, carrying on with the analysis of the principal components of the original price series may not be advisable. Without the stationarity assumptions, it is very likely the case that the usual estimator of the covariance matrix would yields estimates that may be substantially off the mark. Meanwhile, taking the first difference of the logarithm of the price series seems to produce time series that display very little persistence as may be observed from an inspection of Figure 1. Hence, the assumption of stationarity may be more appropriate only after differencing the data. We have substantiated this assumption using the ADF test and the unit root assumption was rejected at the 1% confidence level.

Clearly, taking differences of the log price data entails a loss of information. Nevertheless, an analysis of the differenced data could still yield insight into the factor structure of the original price data since the property could be expected to be shared by both the differenced data and the data in levels. This observation is easily confirmed by means of simple algebraic manipulations. Naturally, the factors that may be extracted from the differenced data would bear very little resemblance to the factors present in the levels data, so that there are limits to how much can be inferred about the data in levels

Table 1: Augmented Dickey-Fuller Tests

(a) 2010

	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.00	2.17	-0.00	-2.16	4.59	-2.86
5 Yr	0.00	2.01	-0.00	-2.00	4.59	-2.86
10 Yr	0.00	2.05	-0.00	-2.05	4.59	-2.86
30 Yr	0.00	2.03	-0.00	-2.02	4.59	-2.86

(b) 2011

	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.00	0.96	-0.00	-0.96	4.59	-2.86
5 Yr	0.00	0.62	-0.00	-0.61	4.59	-2.86
10 Yr	0.00	0.56	-0.00	-0.54	4.59	-2.86
30 Yr	0.00	0.36	-0.00	-0.33	4.59	-2.86

(c) 2012

	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.01	2.51	-0.00	-2.51	4.59	-2.86
5 Yr	0.00	1.31	-0.00	-1.31	4.59	-2.86
10 Yr	0.00	1.22	-0.00	-1.21	4.59	-2.86
30 Yr	0.00	1.36	-0.00	-1.36	4.59	-2.86

(d) 2013

	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.00	1.45	-0.00	-1.45	4.59	-2.86
5 Yr	0.01	1.85	-0.00	-1.85	4.59	-2.86
10 Yr	0.00	1.65	-0.00	-1.65	4.59	-2.86
30 Yr	0.00	1.24	-0.00	-1.25	4.59	-2.86

(e) 2014

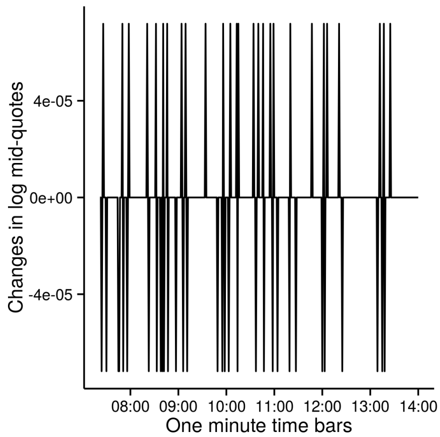
	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.00	1.06	-0.00	-1.05	4.59	-2.86
5 Yr	0.00	1.46	-0.00	-1.46	4.59	-2.86
10 Yr	0.00	1.74	-0.00	-1.73	4.59	-2.86
30 Yr	0.00	1.95	-0.00	-1.93	4.59	-2.86

(f) 2015

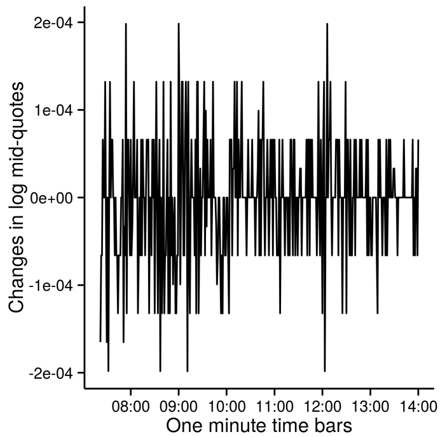
	a	t(a)	lag	t(lag)	5% c.value(a)	5% c.value(lag)
2 Yr	0.01	2.88	-0.00	-2.88	4.59	-2.86
5 Yr	0.01	3.01	-0.00	-3.01	4.59	-2.86
10 Yr	0.01	2.76	-0.00	-2.76	4.59	-2.86
30 Yr	0.00	1.65	-0.00	-1.65	4.59	-2.86

Figure 1: Changes in log prices

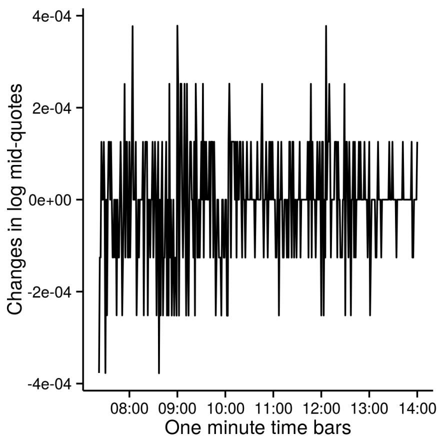
(a) 2 Yr Treasury Note



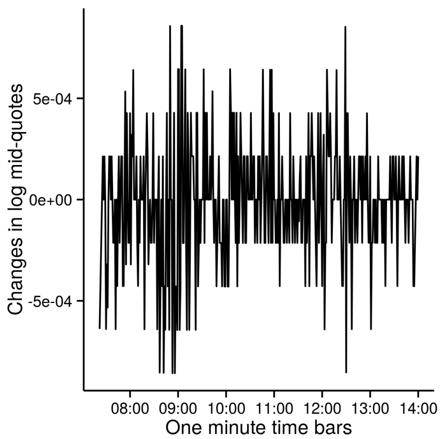
(b) 5 Yr Treasury Note



(c) 10 Yr Treasury Bond



(d) 30 Yr Treasury Bond



once it has been differenced.

Proceeding with the differenced data, we estimate the covariance matrix by means of the unbiased estimator

$$\Sigma := (n-1)^{-1} \sum_{i=1}^{n} x_i x_i',$$

where x is the normalized data series. In the final step we obtain a spectral decomposition of the covariance matrix. That is

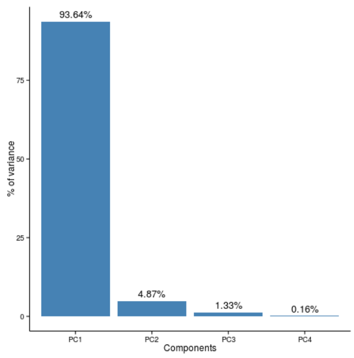
$$\Sigma = \sum_{i=1}^{m} \lambda_i v_i v_i', \tag{3.6}$$

where $\lambda_1^{1/2} \geq \cdots \geq \lambda_m^{1/2}$ are the nonnegative eigenvalues of Σ in descending order of magnitude, and $v_i, i = 1, \ldots, m$, are the corresponding eigenvectors. The eigenvectors are orthonormal, that is they have length one and form a linearly independent set. Hence, from the representation in (3.6), the contribution of the i-th principal factor to the overall variance of the differenced log price data is λ_i. The j-th component of each the i-th eigenvector is the factor loading or beta of the j-th security with respect to the i-th principal component. That is, the components of the eigenvectors summarize exposure levels. For instance, the second element of the third eigenvector is the exposure of the 5 Year Treasury Note Futures contract to the third principal component or risk factor, the so-called curvature factor.

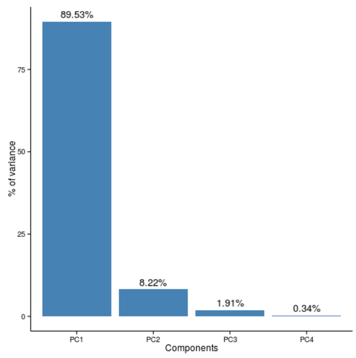
Recall that the eigenvectors are orthonormal so that the associated eigenvalues represent the variance contribution of each principal components to the variance of the differenced log price data. By taking the ratio of individual eigenvalues to the sum of all four eigenvalues, we may estimate the percentage contribution of each principal component to the overall variance. The output of this analysis using subsamples corresponding to each calendar year in our data set is recorded in Figure 2. What is immediately apparent from these figures is that an overwhelming majority of variability is directly attributable to the first component; this component contributes between 90 to 93% of total variability, followed by the second component contributing between 4.8 and 12%. The contributions due to the third and fourth components are fairly modest. The third component contributes between 0.5 and 3%, whereas the fourth component accounts for less than 0.5%. This result is in agreement with previous works such as Litterman & Scheinkman (1991) and Bouchaud et al. (1999) studying the term structure using lower frequency data. By now, this is a stylized fact of the term structure of interest rates, our results confirm this fact for higher frequency data.

Figure 2: Variance contributions

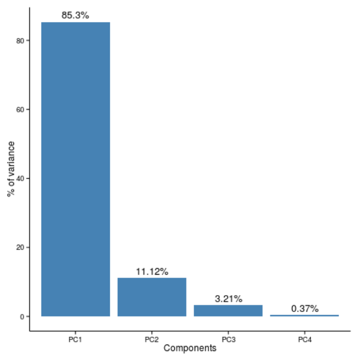
(a) 2010



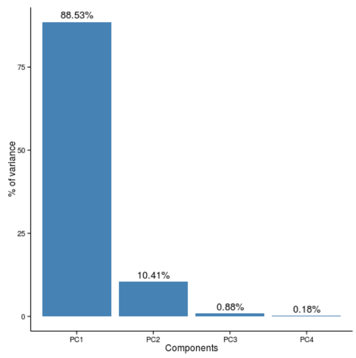
(b) 2011



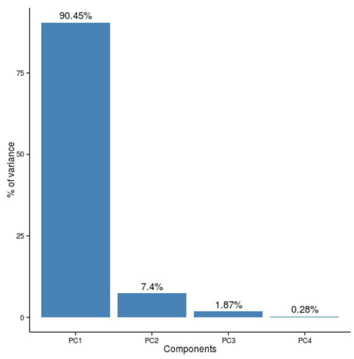
(c) 2012



(d) 2013



(e) 2014



(f) 2015

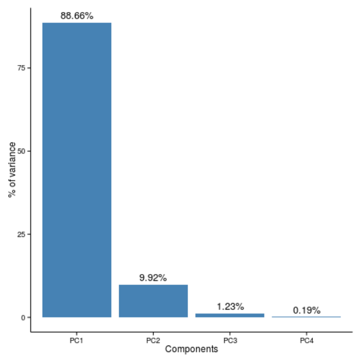
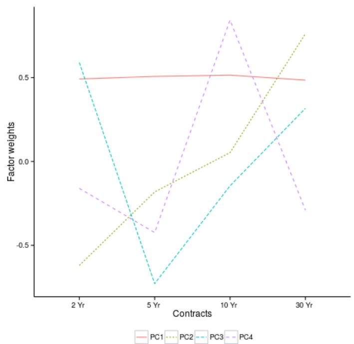
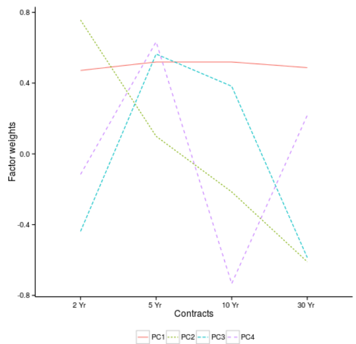


Figure 3: Factor loadings by contract

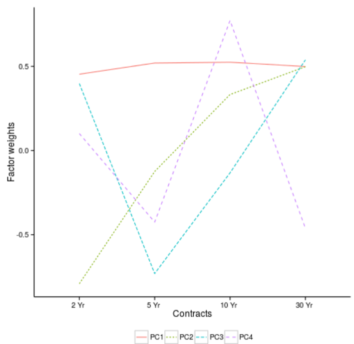
(a) 2010



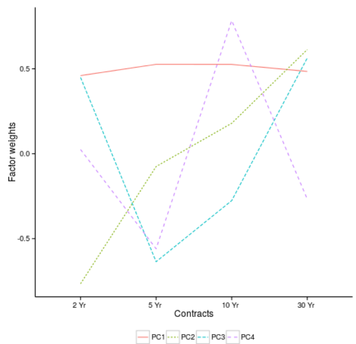
(b) 2011



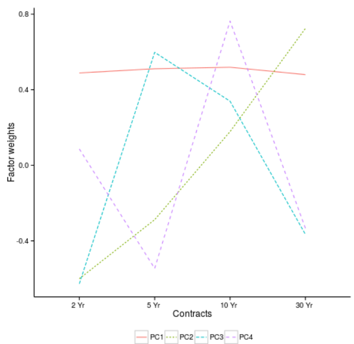
(c) 2012



(d) 2013



(e) 2014



(f) 2015

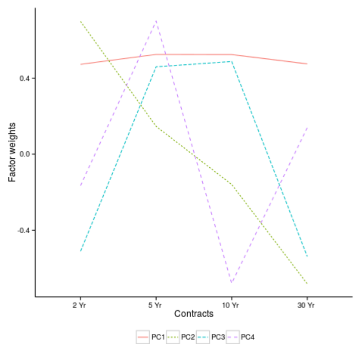


Figure 3, reports the loadings associated with each principal component for a variety of subsamples. The loading for the first fact is fairly stable across maturity and time. The weights are uniformly close to 0.5 so that the effects of shocks emanating from the first factor are felt uniformly across maturities. The loadings associated with the second and third principal components show a great deal of variation across time. In 2010, the effect of a shock emanating from the second factor had the most impact on long bonds than short bonds. The situation was reserved in the following year. A similar reversal may be observed for the third component which in 2010 had a greater impact on medium term bonds than on both long and short bonds.

This empirical analysis of the factors underlying the data forms the basis of the strategy we discuss in the sequel.

3.4 Factor structure implies cointegration

In this subsection we shall study the link between a factor structure description of the yield curve and the existence of cointegrating relationships between contracts of different tenors. An $n \times 1$ vector time series y is cointegrated if each component of y is integrated of order $p > 0$, but there is k, strictly less than n, independent linear combinations of the components of y that result in processes that are integrated of order q, where q is strictly less than p. For our purposes, we shall assume that p is one and q is zero. Hence, cointegration in our setting means that y is a unit root process whose components can be combined linearly in k independent ways to produce stationary processes. The vector of cointegrating relationships are usually normalized and grouped together as the columns of an $n \times k$ matrix denoted β. By definition, β has linearly independent columns, therefore, it has rank $k < n$.

Consider the following model of the yield on n bonds:

$$y_t = Af_t + u_t, \tag{3.7}$$

where y is an $n \times 1$ random vector of yields of varying maturities, A is an $n \times k$ matrix of factor weights, f is a $k \times 1$ random vector of common factors, and u is an $n \times 1$ stationary random vector. Without loss of generality, we may assume that each of the k components of f are unit root processes; otherwise, if only $r < k$ components of f are unit root processes and the remaining $k - r$ are stationary, then we may simply re-write (3.7)

$$y_t = Bh_t + v_t,$$

with $v_t = Cg_t + u_t$, $f_t' = [h_t', g_t']$, and $A' = [B', C']$, where A' is the matrix transpose of A, B and C are, respectively, the $n \times r$ and $n \times (k-r)$ submatrices of A, and h and g are, respectively, r and $k - r$ subvectors of f.

Returning to equation (3.7), the vector of factors may be assumed to be a multivariate random walk, i.e.,

$$f_t = f_{t-1} + \phi(L)\varepsilon_t,$$

where $\phi(L)$ is a lag polynomial, ε is white noise, and $\phi(L)\varepsilon$ is stationary. There some empirical evidence in our data that this assumption is not unjustifiable. Using a matrix analysis argument, details of which may be found in Theorem 3.1 of Escribano & Pena (1993), it is easy to verify that y may be written as a sum of a stationary process and a unit root process:

$$y_t = w_t + z_t,$$

where $w \sim I(1)$ and $z \sim I(0)$. Both w and z may be computed explicitly given the matrix of factor loadings as follows: $w_t = AA'y_t$ and $z_t = (A^{\perp})'A^{\perp}y_t$, where $A^{\perp}$ is the orthogonal complement of A, i.e. $(A^{\perp})'A = 0$. Now, setting $\beta := (A^{\perp})'$, it is easily seen that $\beta y_t = \beta z_t \sim I(0)$, so that β is a matrix of cointegrating vectors. Hence, cointegration of the vector of yields is a consequence of the factor structure of the yield curve.

The analysis in the previous section provides some indirect empirical support for the existence of orthogonal risk factors underlying the dynamics of the term structure. Recall that our analysis of the factors employed differenced price data. So, direct measurements of the risk factors is not an option, but we could at least extract the differenced factors and compute their cumulative sum. While this approach may lack rigor, it nevertheless provides a glimpse of what the original factors might look like. Using the reconstructed risk factors, we test the hypothesis that the level, slope, and curvature factors are unit root processes. The result of this analysis is recorded in Table 2. The results show that the hypothesis of unit root for the risk factors may not be reject at any reasonable level for any of the six calendar years included in our data set.

3.5 Cointegration implies a factor structure

In the previous section, we argued that cointegration is natural assuming the underlying data admits a factor structure. In this section, we argue that the converse is also true. Starting with the assumption that the components of y are integrated of order one, y may be expressed, using lag polynomials, as:

$$(1 - L)y_t = \Phi(L)\varepsilon_t, \tag{3.8}$$

where ε is $n \times 1$ iid noise, L is the lag operator, $\Phi(L) = \sum_{j=1}^{\infty} \Phi_j L^j$, Φ_j is $n \times n$ matrix, and Φ_0 is the $n \times n$ identity matrix. The last condition is an accommodation for the presence of a deterministic linear trend.

Cointegration entails a restriction of the process $\Phi(L)\varepsilon$, and on $\Phi(1)$ in particular. Indeed, writing $\Phi(L) = \Phi(1) + (1 - L)\Phi^*(L)$, where $\Phi^*(L) := (1 - L)^{-1}(\Phi(L) - \Phi(1))$, equation (3.8) may be expressed as:

$$(1 - L)y_t = \Phi(1)\varepsilon_t + (1 - L)\Phi^*(L)\varepsilon_t. \tag{3.9}$$

Solving (3.9) by recursive substitution yields

$$y_t = \Phi(1)z_t + \Phi^*(L)\varepsilon_t, \tag{3.10}$$

where $z_t := \sum_{i=0}^{t-1} \varepsilon_{t-i}$. Now, cointegration implies the existence of an $n \times k$ matrix β, the matrix of cointegrating vectors, such that $\beta' y_t$ is integrated of order 0; but since z_t is a multivariate random walk, it must be the case that $\beta'\Phi(1) = 0$. Now, since β has rank r and Φ spans the subspace orthogonal to its column space, it must be the case that $\Phi(1)$ has rank $n - k$.

Using the Jordan canonical form, we may write

$$\Phi(1) = AJA^{-1}$$

where J is a $(n - k) \times (n - k)$ diagonal matrix containing the non-zero eigenvalues of $\Phi(1)$, A the corresponding $n \times (n - k)$ matrix of eigenvectors, and A^{-1} is the right inverse of A. This decomposition is possible because Φ only has $n - 1$ non-zero eigenvalues. Now setting $u_t := JA^{-1}\varepsilon_t$ and $\nu_t := \Phi^*(L)\varepsilon_t$, and substituting into (3.10) yields

$$y_t = Af_t + \nu_t, \tag{3.11}$$

where $f_t = f_{t-1} + u_t$. The interesting thing about (3.11) is that f is an $(n - k) \times 1$ unit root process driving y. That is cointegration implies a factor structure. This result appears at various levels of generality in Stock & Watson (1988) and Escribano & Pena (1993).

4 Methodology

The basic trading mechanism consists of two main steps. The first step tests for cointegration between the four futures prices and estimates the paramters of a stationary portfolio of the four contracts under the hypothesis of cointegrated prices. The portfolio weights are the components of the cointegration vector. We use a month's worth of daytime (7:30 to 14:00 CT) trading data sampled at one minute intervals for this step. This period is the so-called formation period. Besides estimating the cointegration vector, we also estimate the first two central moments of the stationary portfolio.

Table 2: Augmented Dickey-Fuller Tests

(a) 2010

	Level	Slope	Curvature
intercept	0.00	0.00	0.00
t-stat (intercept)	2.32	2.10	1.87
lag	-0.00	-0.00	-0.00
t-stat(lag)	-2.08	-1.97	-1.88

(b) 2011

	Level	Slope	Curvature
intercept	0.00	-0.00	-0.00
t-stat (intercept)	1.78	-1.74	-1.52
lag	-0.00	-0.00	-0.00
t-stat(lag)	-0.46	-0.31	-0.24

(c) 2012

	Level	Slope	Curvature
intercept	0.00	0.00	0.00
t-stat (intercept)	1.01	1.04	0.85
lag	-0.00	-0.00	-0.00
t-stat(lag)	-1.30	-1.30	-1.50

(d) 2013

	Level	Slope	Curvature
intercept	-0.00	-0.00	-0.00
t-stat (intercept)	-1.26	-1.34	-1.33
lag	-0.00	-0.00	-0.00
t-stat(lag)	-1.51	-1.21	-0.79

(e) 2014

	Level	Slope	Curvature
intercept	0.00	0.00	-0.00
t-stat (intercept)	2.34	2.62	-2.71
lag	-0.00	-0.00	-0.00
t-stat(lag)	-1.75	-2.03	-2.31

(f) 2015

	Level	Slope	Curvature
intercept	0.00	0.00	0.00
t-stat (intercept)	1.89	0.34	1.37
lag	-0.00	-0.00	-0.00
t-stat(lag)	-2.27	-1.56	-1.42

In the second step, we start monitoring prices immediately after the formation period to identify price configurations that may be too rich or too cheap according to our estimates of the first two moments from the formation period. This so-called trading period lasts for about three weeks (100 daytime trading hours) from the end of the formation period. Specifically, we consider the price configuration to present a buy opportunity, if the price of the stationary portfolio falls below two standard deviations of the sample mean computed on the basis of the data generated during the formation period. There is a sell opportuitity if the price climbs beyond two standard deviations of the mean price from the formation period. Hence, a position is entered into whenever the price of the synthetic asset, constructed from the cointegration vector, veers outside the two standard deviation band; the position is long or short according to whether the price configuration is deemed cheap or rich. Short-sale constraints are almost non-existent in the futures market, so they do not enter into our analysis.

Position are opened at any time during the trading period; they are closed as soon as the price of the synthetic asset experiences a large enough correction after its excursion away from the sample mean estimated from the formation period. Specifically, a position is closed as soon as the price falls within the one standard deviation band. Hence, after each correction, at least one standard deviation is earned on the round-trip trade. This process is continued until the end of the trading period at which time all open positions are liquidated at the quoted price. Generally, this is the only time a loss can be registered, since a correction might not have taken place prior to the end of the trading period.

The entire process is repeated on a rolling window from the start of the sample (1 April 2010) to the end of the sample (31 December 2015). In both steps we use exclusively quote data as opposed to transaction data. An advantage of using quote data is that the data is simply more plentiful and may better accurately represent the state of the market as perceived by an agent at any given moment. During the synthetic portfolio formation stage, the cointegration vectors and the first two moments are estimated using the midpoint of the best bid and ask prices. During the trading stage, positions are opened and closed using quoted bid and ask prices: a long position is entered into at the ask and shorts executed at the bid.

The evaluation of the strategy using quotes prices is imperative given the short-term nature of the strategy. All positions are opened for at most 100 daytime trading hours. Theoretically, a position could be entered into and exited the very next minute. For such short investment horizons, the bid-ask spread looms very large. By using quote data, execution costs arising from the bid-ask spread is automatically taken into account. Of course, there are other types of execution costs, but the bid-ask spread is usually the largest source of execution costs, and the use of quoted prices takes care of it right away.

5 Results

5.1 Return calculation

Evaluating the performance of a trading strategy that may involve long and short positions is not altogether a straight foreword matter. In fact, the literature gives little guidance on how to define the one-period *return* of a portfolio consisting of long and short positions. The issue is without complications for a portfolio consisting entirely of long positions; the one-period return is simply the difference between the starting and ending value of the portfolio divided by its starting value. Unfortunately, this definition presents difficulties as soon as portfolios with both long and short positions are considered. For such portfolios, the initial investment could be arbitrarily small, zero, or even negative due to the offsetting effects of long and short positions. In the case of a zero-cost portfolio, the period return is ether positive infinity or negative infinity, regardless of the actual change in the value of the portfolio.

It is easy to see that the standard definition is problematic for portfolios with both long and short positions because the value of the portfolio at the start of the period is always taken as the basis for measuring the performance of the portfolio over the period. By reconsidering the investment simply in terms of cash inflows and outflows much of the difficulties of the standard approach may be overcome. The cash flow perspective, assumes that the entire portfolio is marked to market at the end of each investment period, so that there is a cash flow at the start and end of each period. Cash inflows and outflows are defined from the perspective of the investor. A long position involves an initial cash outflow followed by a cash inflow at the end of the period. The situation is reversed for short positions: an initial cash inflow followed by a cash outflow at the end of the period. Given a portfolio of long and short positions, the one period return is simply the natural logarithm of the ratio of the total cash inflows, from both types of positions, to the total cash outflows, also from both long and short positions. This measure is approximately equal to the ratio of the difference between cash inflows and outflows to cash outflows for the period. That is

$$r_t = \log\left(\frac{\text{Inflows}_t}{\text{Outflows}_t}\right) \approx \frac{\text{Inflows}_t - \text{Outflows}_t}{\text{Outflows}_t}. \tag{5.12}$$

While this definition of period return may seem reasonable for performance measurement in the majority of spot/cash markets, it is not without controversy where futures markets are concerned. Black (1976) observed that it is, in principle, impossible to define fractional or percentage returns for a position in futures contracts. This is because, the time t quoted price of a futures contract is merely the price at which the underlying instrument may be exchanged at an agreed upon future date; no actual transactions occur immediately, so that there are no cash outlays at time t. There is only one transaction, and it occurs at the end of the contract in the form of an outflow or inflow but not both. In practice, both the long and the short sides of a futures contract are required by the trading venue to post collateral to offset the risk of default. Ordinarily, there is a mandated minimum collateral required by the brokerage firm used by the investor. This minimal collateral is otherwise known as the *initial margin*.

A position in futures contracts is marked to market daily, so that favorable price moves results in credits and unfavorable price moves as debits to the margin account. To prevent the margin account from being entirely depleted in the event of a succession of unfavorable price moves, the exchange may set a maintenance margin, which is a minimum balance that must be maintained in the margin account at all times after the initial transaction. Usually, the maintenance margin is the same amount as the initial margin, but it may sometime lower. Margin requirements may differ according to whether the investor is classified as a member of the exchange or a non-member speculator. In 2016, the margin requirements for investors without membership licenses to the Chicago Mercantile Exchange(CME) is 10% more than the margin requirement for members of the CME.

Technically, the margin is not to be taken as an *initial investment*, but it may be argued that it is the amount of cash required to make the transaction possible; without it, the position can not be established. Arguing in this manner, we may define the return of a *long* position in a futures contract at time t to be the change in the price of the contract divided by the initial margin. That is

$$r_t = \frac{F_t - F_{t-1}}{M}, \tag{5.13}$$

where M is the initial margin and F_t is the price at the end of time t of the futures contract. For a short position, the numerator above is multiplied by negative one. This basic definition is also plagued by the usual problems encountered when computing the return generated by a portfolio of both long and short positions. Reasoning as in (5.12), the return metric defined in (5.13) based on the timing of cash flows may be modified to only take into account the direction of cash flows. Hence, given n different futures contracts, we may define the performance metric

$$r_t = \frac{\sum_i^n (\text{Inflows}_{i,t} - \text{Outflow}_{i,t})}{\sum_i \text{Leverage Ratio}_{i,t} \times \text{Par Value}_{i,t} \times Q_{i,t}}. \tag{5.14}$$

where the leverage ratio is simply the ratio of the initial margin of the i-th contract to the par value of the underlying bond, and $Q_{i,t}$ is the exposure, in terms of number of contracts, to the i-th contract at time t. In our setting n is four and the contracts are distinguished by their tenors. Definition (5.12) is a special case of the above; it holds when the position is fully founded, that is, when the leverage ratio is one.

Table 3: Time-averaged CME margin requirements between 1 April 2010 and 12 December 2015.

		Initial margin	
Contracts	Notional value	Members	Speculators
2 Yr	200000	448	493
5 Yr	100000	818	900
10 Yr	100000	1323	1456
30 Yr	100000	2647	2912

The initial margins are ordinarily not the same across contracts and, therefore, must be handled carefully. For instance, in the last quarter of 2016, the initial margin for the 30-Year Treasury Bond Futures contract was \$4000, whereas the initial margin of the 2-Year Treasury Note Futures contract was only \$550. Beside the differences in initial margins by contract types, there are also variations over time. For most of 2010, the initial margin requirement for the 5-Year Treasury Note Futures contract was \$800 for investors with membership licenses and \$880 for non-members. Meanwhile, for all of 2014, the initial margin for the same contract was \$900. To simplify our analysis, we compute a time-weighted average of the initial margin for the time period between 1 April 2010 and 31 December 2015 for each contract type. The time weighted average for members and non-members of the CME are recorded in Table 3. As may be expected, margin requirements increase with the tenor of the underlying, because the price of contracts with longer maturities are more likely to experience large price swings.

As previously stated, the initial margin is merely the minimum collateral required to initiate a transaction in one futures contract. An investor may chose to apply however much collateral he or she desires. If each transaction is fully funded, i.e., if the exact amount of the exposure to each contract is always set aside for each transaction, then the appropriate performance measure would be a slight modification of the formula given in (5.12). That is

$$r_t = \frac{\sum_i^n (\text{Inflows}_{i,t} - \text{Outflow}_{i,t})}{\sum_i \text{Outflow}_{i,t}}. \tag{5.15}$$

We remark that the use of fully funded accounts in the treasury futures market is not very common. Consider that the notional value of the 2-Year treasury futures contract is \$200,000 and \$100,000 for the others. Hence, putting together a portfolio consisting of even a small number of contracts quickly becomes prohibitively capital intensive. Meanwhile, treasury futures, even those written on long bonds, have relatively stable long-term prices. As a result, investments in the treasury futures market are most often undertaken using leverage or a margin account.

We conclude this section with a remark on the distinction between an investment period and a trading period. Trading periods are fixed: they are exactly 6000 daytime trading minutes, approximately 14 trading days. An investment period is simply the time between when a position is opened and the time when it is closed. Positions are opened when the price configurations of the four securities indicate a departure from the stable relationship established during the preceding formation period. The positions are closed when the stable relationship is restored. This deviation and restoration towards a stable relationship may occur several times during a single trading period, thereby creating multiple opportunities and, hence, investment periods.

The return formulas in (5.14) and (5.15) relate to the return over a single investment period. For trading periods with multiple periods, we compute the return over the trading period as the sum of the individual returns generated from each investment period contained within the trading period. That is

$$r_t = \sum_{i=1}^{q} r_{i,t}$$

where $r_{i,t}$ is the return, computed via formula (5.14) or (5.15), of the i-th investment period of the t-th trading period, and q is the total number of investment periods occurring in the t-th trading period.

5.2 Excess returns

We summarize the distribution of returns generated by backtesting the cointegration strategy described in the previous section in Table 4. The backtest

Table 4: Annualized (100 trading hours) returns on initial margin and fully funded account.

Panel A: Fully-funded excess return over the equal-weighted portfolio

	Ratio	Difference
Average return	0.0604	0.05995
Standard error (Newey-West)	0.02625	0.02737
t-Statistic	2.3012	2.18984
Excess return distribution		
Median	0.04749	0.03655
Standard deviation	0.30453	0.35263
Skewness	2.12072	1.66976
Kurtosis	14.58188	9.49711
Minimum	-0.68255	-0.75226
5% Quantile	-0.29139	-0.35234
95% Quantile	0.51928	0.61858
Maximum	1.85812	1.79468
% of negative excess returns	40.625	44.79167

Panel B: Return on margin account

	Ratio	Difference
Average return	15.12327	14.95951
Standard error (Newey-West)	3.87113	4.13448
t-Statistic	3.90668	3.61823
Excess return distribution		
Median	0.6238	0
Standard deviation	46.01987	51.17213
Skewness	1.63053	1.29959
Kurtosis	10.99074	7.87337
Minimum	-105.68033	-110.64147
5% Quantile	-51.1802	-70.32086
95% Quantile	83.25456	96.94193
Maximum	259.66658	250.56902
% of negative excess returns	14.58333	16.66667

is run over daytime trading hours between April 1, 2010 and December 31, 2015. The entire period is divided into 96 hundred-hour trading periods lasting approximately three business weeks. The figures shown in the table are the annualized hundred-hour returns. We show cash flow-based returns computed on a fully funded account in the first panel of the table; the second panel displays the distribution of cash flows-based returns computed using the initial margin as the cost basis. Each panel reports two sets of backtest results: one for prices adjusted backwards by the application of a proportional factor (Ratio) and the other for prices shifted in levels backward by the amount of the roll-induced price gap (Difference).

Now, the annualized excess return over the equally weighted portfolio of all four contracts, assuming a fully-funded account, are 6.01% and 6.00% respectively for the ratio and the difference price adjustment procedures. The Newey-West adjusted t statistics are 2.3 and 2.2, respectively. Given this result, the hypothesis that the cointegration strategy dominates the equal-weighted portfolio, cannot be rejected. The idea is that one may short as many of the equal-weighted portfolio as necessary and use the proceeds to set up the cointegration strategy without incurring a loss.

Meanwhile, the annualized return using the initial margin as the cost basis are, respectively, 1500% and 1490%. These returns are not as preposterous as they first seem. Consider that the leverage factor implicit in the initial margin for the 2-Year contract is 446 and that of the 5-Year contract is 122. The inflated returns are, therefore, merely a consequence of the inflated leverage factors. The t statistics in both cases are in excess of 2. Also, note that the out-sized returns that may be achieved by trading on the initial margin come at the expense of taking significant risks: consider that the standard deviation of the returns on initial margin are 159.17 times the volatility of the return on the fully funded account. Clearly, in practice, what an investor ends up doing would be somewhere between trading a fully-funded account and posting the minimum required collateral. At any rate, our analysis provides a starting point for reasoning about how to incorporate leverage in a more realistic real world strategy.

5.3 Statistical Arbitrage

Stephen Ross (1976) gave the first serious treatment of the concept of arbitrage. While his treatment might have been of a heuristic nature, it nevertheless conveyed the essence of an arbitrage, which is a trading strategy that yields a positive payoff with little to no downside risk. The first rigorous definition appeared in Huberman (1982), where it was defined in the context of an economy with asset generated by a set of risk factors as a sequence of portfolios with payoffs ϕ_n such that $E(\phi_n)$ tends to $+\infty$ while $\mathrm{Var}(\phi_n)$ tends to 0. The concept has undergone numerous changes in the literature, see for example Kabanov (1996), but the essential meaning of the term as a low risk risk investment still remains. We focus on a special type of arbitrage known in the empirical asset pricing literature as a *statistical arbitrage*, which Hogan et al. (2004) defines as a zero-cost, self-financing strategy whose cumulative discounted value $v(t)$ satisfies:

1. $v(0) = 0$,
2. $\lim_{t\to\infty} E(v_t) > 0$,
3. $\lim_{t\to\infty} P(v_t < 0) = 0$, and
4. if $v(t)$ can become negative with positive probability, then $\mathrm{Var}(v_t < 0)/t \to 0$ as $t \to \infty$.

Even though the above notion of arbitrage bears resemblance to the original definition given by Huberman (1982), it is worth noting that there is a crucial difference between the two concepts. In the first instance, the limit is taken with respect to the cross-section of the economy, i.e. the sequence of small economies is assumed to expand without bound, whereas the definition given above requires the investment horizon to tend to infinity. It is easily verified that the first three conditions, assuming the existence of the first moment correspond to the definition of an arbitrage in the classical sense of (Delbaen & Schachermayer, 1994). Chapter 2 of Dare (2017), obtains a series of equivalent characterizations of market efficiency. In particular, Dare (2017, Proposition 2.1) shows that a necessary condition for market efficiency is the existence of a local martingale measure after expressing asset prices in units of any strictly positive convex portfolio of the zeroth asset and any other asset (possibly itself). To apply this result, we express prices in units of the 2-Year treasury futures contract and attempt to exhibit a violation of the no-arbitrage condition.

Following Hogan et al. (2004) we propose to test market efficiency under the assumption that the change in the discounted cumulative gains of the strategy satisfies:

$$v_t - v_{t-1} = \mu + \sigma t^{\lambda} z_t, \tag{5.16}$$

where t is an integer; $\sigma, \lambda,$ and μ are real numbers; and z_t is an i.i.d. sequence of standard normal random variables. The model allows for deterministic time variation in the second moment, but makes the seemingly strong assumption that there is no serial correlation between returns. Our own simulations reveal that the effects of serial correlations are slight. In fact, assuming z were generated by an AR(1), differences of more than 10% on average standard errors only start to occur for values of the autoregressive parameter in excess of 0.9 in absolute value. Since, the sample autocorrelation of the returns of the strategy is only -0.158, it is likely the case that serial autocorrelation is a minor issue.

The inference strategy we have adopted is not without weaknesses. A stylized empirical fact of financial markets is that asset returns generally have fat-tail distributions. The normality assumption may therefore seem overly restrictive. Moreover, the adopted parametric model may itself be a source of misspecification errors. A more sophisticated analysis would perhaps employ robust tools such as the bootstrap.

While the above criticisms may be valid, note that the test statistics discussed in Hogan et al. (2004) and Dare (2017, Chapter 2) are very conservative because they rely on the Bonferroni criterion, which stipulates that in compound tests involving a joint-hypothesis, the sum of the p-values of the individual tests is an upper limit for the Type I error of the compound test (Casella & Berger, 1990, p.11).

The test of efficiency then consists of estimating the parameters of (5.16) and then testing for

1. $\mu > 0$, and
2. $\lambda < 0$.

For, under the null hypothesis, the cumulative returns will increase without bound as the variance of the cumulative gain tends to zero. Under the assumption of normally distributed z_t, the test can be carried out by maximum likelihood estimation. Table 5 summarizes the results of estimating model (5.16). The p-value of the joint test using the Bonferonni correction, which consists of adding up the p-values from each sub-hypothesis, of the presence of a statistical arbitrage is approximately 0.02 in both the ratio and difference-adjusted price series.

Table 5: Test for statistical arbitrage

	Difference			Ratio		
Par.	Estimate	Std Error	p-value	Estimate	Std Error	p-value
μ	0.3351	0.0108	<0.01	0.3735	0.0106	<0.01
σ	1.151	0.5397	0.041	4.98	3.2077	0.1195
λ	-0.6808	0.1288	<0.01	-1.0353	0.1779	<0.01

The p-value for the estimates of σ are relatively large, estecially for the ratio adjusted price series. This is to be extected since for large t, the volatility of incremental payoffs is mostly determined by the term t^{λ} and not σ. We conduct a separate test with σ restricted to 1. The output of that test is recorded in Table 6. The estimates and p-value obtained from the restricted model match the estimates from the unrestricted model. These results provide a strong indication that the strategy not only earns positive excess return over the equal weighted portfolios, but also, a positive free-lunch.

Table 6: Test for statistical arbitrage ($\sigma = 1$)

	Difference			Ratio		
Par.	Estimate	Std Error	p-value	Estimate	Std Error	p-value
μ	0.333	0.0082	<0.01	0.3474	0.0104	<0.01
λ	-0.6452	0.0199	<0.01	-0.5885	0.0204	<0.01

6 Conclusion

Starting with the assumption that interest rates and, therefore, bond futures prices admit a factor structure, we evaluate a trading strategy based on the assumption of cointegrated bond futures prices. We argue that coitegration is natural if in fact the dynamics of the yield curve is driven by orthogonal risk factors which together form a jointly unit root process. Direct verification of this hypothesis is difficult because the price series and its log transfomed counterparts are likely not stationary. On the other, the stationary assumption can be made and tested using the change in log prices. Using differenced price data, we argued empirically that the vast majority of the volatility experienced by the changes in log prices may arise from three dominant risk factors. Since the factors are othogonal and the differenced log prices may be assumed to be stationary, there is very little doubt that the factors contained in the differenced log prices are stationary. To test the claim for the log prices in evels, we estimated the factors in levels by taking cumulative sums of the factors estimated using differenced log prices. For these proxies of the factors in levels, the assumption of unit roots could not be rejected at reasonable significant levels.

With the choice of cointegrating strategy properly motivated, we proceeded to evaluate a simple trading strategy based on the cointegration hypothesis. The crust of the strategy consists in opening a position as soon the price configuration appears to deviate from an estimated stable cointegration relationship. The strategy is evaluated by computing the ratio of cash-equivalent inflows to outflows. We also consider a return metric based on the initial margin required to take either a short or a long position in one futures contract. Our results reveal that the gains from this strategy are both economically and statistically significant. This exercise allows to argue along the lines innitiated by Jarrow & Larsson (2012) that the U.S. treasury bond futures market, for the period for which we have data, was not informationally efficient.

References

Black, Fischer (1976) "The pricing of commodity contracts", *Journal of Financial Economics*, Vol. 3, pp. 167–179.

Bouchaud, Jean-Philippe, Sagna, Nicolas, Cont, Rama, El-Karoui, Nicole, and Potters, Marc (1999) "Phenomenology of the interest rate curve", *Applied Mathematical Finance*, Vol. 6, pp. 209–232.

Bradley, Michael G. and Lumpkin, Stephen A. (1992) "The treasury yield curve as a cointegrated system", *The Journal of Financial and Quantitative Analysis*, Vol. 27, No. 3, pp. 449–463.

Casella, George and Berger, Roger L. (1990) *Statistical inference*: Thomson Learning, 2nd edition.

Christensen, J.H.E, Diebold, F.X., and Rudebusch, G.D (2011) "The affine arbitrage-free class of nelson-siegel term structure models", *Journal of Econometrics*, Vol. 164, pp. 4–20.

Cox, John C., Jonathan E. Ingersoll, Jr., and Ross, Stephen A. (1981) "The relation between forward prices and futures prices", *Journal of Financial Economics*, Vol. 9, pp. 321–346.

Dare, Wale (2017) "On market efficiency and volatility estimation", Ph.D. dissertation, University of St.Gallen, Switzerland.

Delbaen, Freddy and Schachermayer, Walter (1994) "A general version of the fundamental theorem of asset pricing", *Mathematische Annalen*, No. 300, pp. 463 – 520.

Diebold, Francis X. and Li, Canlin (2006) "Forecasting the term structure of government bond yield", *Journal of Econometrics*, Vol. 130, pp. 337–364.

Diebold, Francis X. and Rudebusch, Glenn D. (2013) *Yield Curve Modelling and Forecasting*: Princeton University Press.

Duffie, Darrell and Kan, Rui (1996) "A yield-factor model of interest rates", *Mathematical Finance*, Vol. 6, No. 4, pp. 379–406.

Dybvig, Philip H., Jonathan E. Ingersoll, Jr., and Ross, Stephen A. (1996) "Long forward and zero-coupon rates can never fall", *The Journal of Business*, Vol. 69, No. 1, pp. 1–25.

Escribano, Alvaro and Pena, Daniel (1993) "Cointegration and common factors", Working paper 93-11, Universidad Carlos III de Madrid.

Fama, Eugene (1969) "Efficient capital markets: A review of theory and empirical work", *The Journal of Finance*, Vol. 25, No. 2, pp. 387–417.

Filipović, Damir (1999) "A note on the nelson-siegel family", *Mathematical Finance*, Vol. 9, pp. 349–359.

Hall, Anthony D., Anderson, Heather M., and Granger, Clive W. (1992) "A cointegration analysis of treasury bill yields", *The Review of Economics and Statistics*, Vol. 74, No. 1, pp. 116–126.

Hogan, Steve, Jarrow, Robert, Teo, Melvyn, and Warachka, Mitch (2004) "Testing market efficiency using statistical arbitrage with applications to momentum and value strategies", *Journal of Financial Economics*, Vol. 73, pp. 525 – 565.

Hubalek, Friedrich, Klein, Irene, and Teichmann, Josef (2002) "A general proof of the Dybvig-Ingersoll-Ross theorem: Long forward rates can never fall", *Mathematical Finance*, Vol. 12, No. 4, pp. 447–451.

Huberman, Gur (1982) "A simple approach to Arbitrage Pricing Theory", *Journal of Economic Theory*, Vol. 28, No. 1, pp. 183–191.

Jarrow, Robert A. and Larsson, Martin (2012) "The meaning of market efficiency", *Mathematical Finance*, Vol. 22, No. 1, pp. 1–30.

Kabanov, Y.M. (1996) "On the FTAP of Kreps-Delbaen-Schachermayer", *Statistics and Control of Stochastic Processes*, pp. 191–203.

Labuszewski, John W., Kamradt, Michael, and Gibbs, David (2014) "Understanding treasury futures", report, CME Group.

Litterman, Robert B. and Scheinkman, José (1991) "Common factors affecting bond market returns", *The Journal of Fixed Income*, Vol. 1, No. 1, pp. 54–61.

Merton, Robert C. (1973) "Theory of rational option pricing", *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1, pp. 141–183.

Nelson, Charles R. and Siegel, Andrew F. (1987) "Parsimonious modeling of yield curves", *The Journal of Business*, Vol. 60, No. 4, pp. 473–489.

Ross, Stephen A. (1976) "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, Vol. 13, pp. 341–360.

Singleton, Kenneth J. (2006) *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*: Princeton University Press.

Stock, James H. and Watson, Mark W. (1988) "Testing for common trends", *Journal of the American Statistical Association*, Vol. 83, No. 404, pp. 1097–1107.

Zhang, Hua (1993) "Treasury yield curves and cointegration", *Applied Economics*, Vol. 25, No. 3, pp. 361–367.