The Term Structure of Interest Rates, Unit Roots and Cointegration

Topics in Statistics and Econometrics
by Professor Dr. Joachim Grammig
Supervision: Dr. Silika Prohl

Summer Term 2007

Markus Niedergesäss
Ruth-Marx-Strasse 15
72072 Tübingen
International Economics
9$^{\underline{th}}$ *Semester*

Contents

List of Tables

1 Introduction

Term structure of yields associated with bonds of different maturity was heavily discussed in recent years, since the idea of cointegration was brought up by Engle and Granger(1987). Ever before, the empirical analysis of the relationship between yields, which typically move together, was quite difficult, because interest rates are assumed to be rather I(1) than stationary processes. Cointegration now made it possible to link the movements of interest rates with different maturities in long run equilibriums. Then error correction models show how interest rates react on deviations from those equilibriums.

Fundamental literature on this topic was written by Campbell and Shiller(1987) and particularly Hall, Anderson and Granger(1992), who extended their approach from a bivariate to a multivariate case. This paper orientates on that by Hall et al.. However, we now are able to observe data from 1954:07 to 2007:03. We also divided in two subsamples, due to a regime shift in 1979. For the first subsample the results remain the same. Nevertheless, the second subsample can not support theory with the idea, that the term structure of interest rates is well modeled as a cointegrated system. We will have a closer look on that.

The reminder of this paper is as follows. Section II sets a theoretical framework, which relates cointegration and error correction models to term structure models. Section III presents the data used and mentions some problems. Section IV presents the test methods applied and the results obtained. We also provide a discussion on the results and refer to problems with the test procedures. Then Section V gives an short overview about the results from an error correction model, which was implemented for the first subsample. Section VI will conclude.

2 Theoretical Framework

In the following I will introduce a theoretical framework, which likes to show possible cointegration of interest rates with different maturities. The framework bases on that by Hall et al.(1992), which is also used in many other papers, such as Lanne(2000) for example. We let $r(k,t)$ be a k-period interest rate, while $f(k,t)$ is a forward interest rate for an one period pure discount bond maturing at time $t+k$. Then, with $f(1,t) = r(1,t)$ we obtain by means of the Fisher-Hicks formulae,

$$r(k,t) = \frac{1}{k}\left[\sum_{j=1}^{k} f(j,t)\right] \text{ for } \quad k = 1,2,3,\ldots \quad . \tag{1}$$

It states that the k-period interest rate $r(k,t)$ can be calculated as a sample mean of the k forward rates $f(j,t)$.

The forward rate $f(j,t)$ is equal to an expectation on $r(1, t+j-1)$, that is the yield of a one-period discount bond, which matures at time $t+j$. Therefore, it is possible to assume the following relation:

$$f(j,t) = E_t\left[r(1, t+j-1)\right] + \Lambda(j,t), \tag{2}$$

where E_t denotes the expectation on the yield of a one-period discount bond at time $t+j-1$ and $\Lambda(j,t)$ is a premia, that can be due to risk aversion or deviations in investors´s preferences about liquidity, while for a technical analysis it can be seen as a disturbance term.

By substitution of equation (2) into equation (1) we receive a general relationship between interest rates with different maturities,

$$\begin{aligned} r(k,t) &= \frac{1}{k}\left[\sum_{j=1}^{k} E_t\left(r(1, t+j-1)\right)\right] + \frac{1}{k}\left[\sum_{j=1}^{k}\Lambda(j,t)\right] \\ &= \frac{1}{k}\left[\sum_{j=1}^{k} E_t\left(r(1, t+j-1)\right)\right] + L(k,t). \end{aligned} \tag{3}$$

Equation (3) implies that the k interest rates with different maturities are tied in one long-run equilibrium and will move together. There are different assumptions about the premia $L(k,t)$. One is the pure expectations hypothesis, which assumes $L(k,t) = 0$. An other assumption only requires the premia to be constant over time, so that the mean might differ from zero. For later estimations stationarity is required.

As said before, by means of equation(3) we have derived a general relationship, however we can not use it directly for further empirical studies. This is in particular due to non-stationarity of interest rates. Interest are rather assumed to be $I(1)$-processes and there is considerable evidence, that supports this assumption. As a consequence we have to bring in the idea of cointegration, that I will explain briefly.

Cointegration The idea of cointegration was first mentioned by Engle and Granger(1987). For the following only cointgration of $I(1)$-processes will be recognized. Now, let $Z(t)$ contain two such $I(1)$-processes. If there exists a parameter a such that the linear combination $a'Z(t)$ is stationary, than the two $I(1)$-series are cointegrated and a is called a cointegrating vector. However, one has to be cautious, because a is not a unique solution. There always exists a nonzero scalar b so that ba is again a cointegrating vector.

For any k-period interest rate a possible cointegration with the one period interest rate can be shown by rearranging equation (3). We finally obtain[1]

$$[r(k,t) - r(1,t)] = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{j=i} E_t \Delta r(1, t+j) + L(k,t), \tag{4}$$

where $\Delta r(v,l) = r(v,l) - r(v,l-1)$. If the right hand side of equation(4) is stationary, that is if $\Delta r(1,t)$ and $L(k,t)$ are stationary, then one possbile cointegrating vector would be $\beta = [1, -1]$. This result implies that any of the k yields of $r(k,t)$ is cointegrated with the one-period interest rate at time t. Defining the spreads between two yields as $s(i,j,t) = r(i,t) - r(j,t)$, cointegration also implies that those spreads are stationary linear combinations of the vector series $X(t)$.

Of course, cointegration is also possible for more than just two time-series. Since, any k-period interest rate is cointegrated with the one-period interest rate, we can assume a new vector series $X(t) = [r(1,t), r(k_2,t), \dots, r(k_n,t)]$, that contains n different yield series, including the one-period interest rate, where $k_2, \dots, k_n$ are the maturities for the $n-1$ other yield series. Equation (4) has shown that any interest rate with maturity $k_2, \dots, k_n$ is cointegrated with the one period interest rate. Thus there exists a $(n-1) \times n$-matrix A of the following form

$$A = \begin{pmatrix} -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

for which the linear combination $A'X(t)$ is stationary. It is easy to see that the columns a_i of A are all linearly independent. That means there does not exist a scalar b so that i.e. $a_1 = ba_2$. However, as seen in the two variable case before, the vectors a_i are not unique, there always exists a vector h, such that the scalar $h'AX(t)'$ is stationary. This result is quite important for our case of interest rates, as it implies, that any spread of two interest rates with different maturities k_i is stationary. Hence those two yields are cointegrating.

Since all vectors a_i are linearly independent the rank of A is equal to $n-1$. Then also the cointgrating space has rank $n-1$ and the rows $a_1, \dots, a_{n-1}$ form a basis for this cointegrating space. Stock and Watson(1988) found out that there exist k common trends for system of n variables and a cointegration rank, that equals $n-k$. For our case of interest rates with a cointegration rank of $n-1$, there has to exist one common trend. We can therefore write

[1] for further details on how to rearrange equation (3) see Appendix

our system of interest rates as follows

$$\begin{array}{lcl} r(1,t) & = & u_1 + c_1 e(t) \\ r(2,t) & = & u_2 + c_2 e(t) \\ \vdots & & \vdots \\ r(k,t) & = & u_k + c_k e(t) \end{array} \quad ,$$

where u_j is a stationary component, c_j is a parameter and $e(t)$ is the common stochastic trend. This common trend drives the time series behavior of every interest rates and since the other components are all stationary, this common trend dominates the behavior of each yield.

A further implication of the finding that any spread is cointegrating, is that all the results that we obtained so far, have to hold theoretically for any set of $n-1$ linearly independent spread vectors.

Vector Error Correction Model Cointegration implies and is implied itself by an error correction representation, as shown by Engle and Granger(1987). A VECM can be derived from a typical VAR with p-lags

$$X(t) = \Phi_1 X(t-1) + \Phi_2 X(t-2) + \ldots + \Phi_p X(t-p) + \varepsilon(t) \quad , \tag{5}$$

where for our purposes $X(t)$ is the vector series from above, and $\varepsilon(t)$ is a vector of white noise disturbances, which may be contemporaneously correlated. The resulting VECM is then of the following form

$$\Delta X(t) = \pi X(t-1) + \sum_{i=1}^{p-1} \pi_i \Delta X(t-i) + \varepsilon(t) \quad , \tag{6}$$

where $\pi = -(I - \sum_{i=1}^{p} \Phi_i)$ and $\pi_i = -\sum_{j=i+1}^{p} \Phi_j$.

Now, I will focus on the $n \times (n-1)$ matrix π, which is can be written in the form $\pi = \alpha\beta'$, where β is a cointegrating vector for $X(t)$. For economic purposes it is now possible to consider an underlying equilibrium relationship $\beta' X(t) = \mu$, where μ is vector containing the equilibrium values. We may also assume that agents react on a disequilibrium error $\beta' X(t) - \mu \neq 0$. The vector α can then be seen as a matrix of adjustment coefficients, displaying the speed with which the variables return to the equilibrium. However, this assumption has to be treated cautiously, I will explain it more detailed soon. In our case of interest rates with different maturities the cointegrating vector β equals our matrix A from above. This result implies that any spread between two yields is tied in one long-run-equilibrium. We may now rewrite the above VECM to receive

$$\Delta X(t) = \alpha^* \left[s(t-1) - \mu \right] + \pi_i \Delta X(t-i) + \varepsilon(t) \quad , \tag{7}$$

where $[s(t-1) - \mu]$ is the error correction term and α^* is a matrix containing the adjustment coefficients. The spread-vector $s(t-1)$ is obtained as a linear combinations of $A'X(t)$, so that it may contain every possible spread between two yields, while as a matter of course α^* has to be modified for different spreads.

Returning to equation (6) there is another important fact we should be aware of. We have already seen that the rank of A equals the rank of the cointegrating space. Since π is a linear combination of A, the rank of π equals that of A, if α has full rank. Therefore, we have now the possibility to obtain the rank of the cointegrating space by checking for the rank of π. For $rk(\pi) = 0$ the VECM is equal to a VAR estimated in first differences, while for a $1 \leq rk(\pi) \leq (n-1)$ we have $n-1$ linearly independent cointegrating relationships, what implies cointegration between any yields of different maturities. Testing for statistical significance of α^* in equation (7) is a similar way for checking for cointegration.

One point already mentioned also has to be regarded. As said, the assumption that agents react on a disequilibrium error has to be treated cautiously. Hall et al. (1992) mention that a VECM does not necessarily imply an adjustment of yields because spreads are out of equilibrium. It could be rather the point as shown by Campbell and Shiller (1987, 1988) that spreads might measure anticipated changes in yields. This is based on the idea that agents have more information for forecasting changes in yields than the history of short yields alone.

3 Data

The data used in this study consist of monthly nominal US Treasury bill rates from the Board of Governors of the Federal Reserve System in Washington. The file contains ten series for Treasury bills with different maturities, that goes from overnight interest rates up to yields with thirty years to maturity. However, only five series are available for the full distance from 1954:7 to 2007:3. Availability is provided for the overnight rate and the yields with one, three, five and ten years to maturity. The overnight rate is used because data is available for a long period. However, it has to be treated carefully since it has not a maturity of $k = 1$, although it is treated as it would, since data for the yield with one month to maturity is only available from 2001 up to now.

The full sample are separated in two subsamples which go from 1954:07 to 1979:09 and from 1979:10 to 2007:03. This was done due to a regime change in the degree of interest rate targeting undertaken by the Federal Reserve and is line with other papers like Hall et al.(1992) and Lanne(2000). Nevertheless, the second subsample contains a lot more data now, then the two papers mentioned. Regime shifts are structural breaks, that may lead to distortions of our empirical results, when we test on whether the series contain unit or whether they cointegrated.

For the time span of 1954:07 to 1979:09 the interest rate was fully targeted, while there was only a partial targeting for the latter subsample. Hall et al. divide their latter subsample a further time, due to a second regime shift in 1982. Because it was a slight one, I will follow Lanne who only appoints the change in 1979.

As already mentioned monthly data is used in this paper. This is by virtue of a lack of data, if the frequency would be increased on weekly or even daily basis. However, it is worth noticing that a latter study from Choi and Chung(1995) has shown that an increase in frequency may significantly improve the finite sample power of the Augmented Dickey Fuller test for data like interest rates. Moreover, an increase in the time span also increases the power of unit root tests.

4 Econometric Methodology and Empirical Results

This section will give an overview from testing the implications due to the theoretical framework. First the tests, which have been applied are introduced, then the results are presented. Finally the results are discussed and several problems linked with the tests, that are presented.

4.1 Unit Root Tests

First of all I will have a look on the unit root assumption of interest rates, since it is crucial for all further results theoretically, as well as empirically. Problems with the assumption of interest rates as realizations of unit root processes mainly arise due to the fact that interest rates are bounded, i.e. they can not take values lower than zero, at least nominally. Hence, interest rates can not be seen as exact unit roots. Moreover we should ask if the assumption can be taken for granted over the long run. As John Cochrane remarked: ”Interest rates now are the same as in Babylonian days. How can there be a unit root in interest rates?” The assumption of interest rates being realizations of unit root processes may thus only hold for certain time spans. However, testing this point becomes quite difficult, since frequent data is only available for a short time span of maybe a hundred years. It may also be the point, that the unit root assumption holds for time spans referring to certain regimes. Then tools like Markov-Regime-Switching-Models come into play, though they will not be used here.

Augmented Dickey-Fuller Test I applied two test procedures for testing on unit roots, one is the Augmented Dickey Fuller test(ADF). The original Dickey-Fuller test assumed the errors to be independently and identically distributed. The ADF test allows for serially correlation, therefore difference lags are included in the test equation, which looks as

$$\Delta r(j,t) = (\rho - 1)r(j,t-1) + \theta_1 \Delta r(j,t-1) + \ldots + \theta_k \Delta r(j,t-k) + \epsilon(t) \quad , \tag{8}$$

where the lag length k is chosen in order to ensure that $\epsilon(t)$ are purely random. The lag length was selected by the Schwartz Information Criterion(SIC). The ADF test tests the null hypothesis that the series contains a unit root, that is

$$H_0: \quad \rho = 1 \quad \text{vs.} \quad H_1: \quad \rho < 1 \quad .$$

The asymptotic test distribution is not a χ^2-distribution, it rather depends on whether trends or constants are used[2].

Dickey-Fuller GLS test The Dickey-Fuller GLS test is quite similar to the ADF test, however, the time series is detrended before it is tested on a unit root. The null hypothesis is the same as with the ADF test and lag length is also selected via SIC. The test distribution depends also on how the test is implemented.

The ADF test is used, although Maddala et al.(1998) considers it as useless for practical purposes, since it has low power problems. However, it is a standard test and was also used in Hall et al. I also like to refer some words to why only a constant is used in the test statistics. As Hamilton(1994) says, there is no economic reason why time series of interest rates should contain a linear trend, since they do not show any behavior that supports this idea. We are also in line with Lanne(2000) and Hall et al.(1992).

Results Table 1 reports the results from the ADF and the DF-GLS tests. On the 5% significance level both tests can not reject the H_0, that the series contains a unit root, except for the overnight rate of the second subsample. In this case, howver, only the ADF test can reject the H_0, while the DF GLS test can not reject. The assumption of the interest rates being realizations of unit root processes can not be rejected for both subsamples, as for the full sample as such. The lag length selected by the SIC is also reported in Table 1. The lag length differs strongly between different interest rates, but does not differ between the two tests except for the case of the overnight rate of the second subsample. Here fifteen lags are recommended for the DF-GLS test, while fourteen are used for the ADF test.

It is worth to have a closer look on the results. For any sample the probability of non-

[2] Asymptotic test distributions for the ADF test can be seen in the Appendix, Critical Values for the DF-GLS test can e.g. be obtained from Elliott et al.(1996)

Table 1: Unit root tests for the full sample and the two subsamples

Interest rate	lag length	ADF t-statistics	prob.	DF-GLS t-statistics
1954 : 07 - 2007 : 03				
r^1	13	-2.759	0.0649	-1.407
r^{12}	6	-2.167	0.2185	-0.863
r^{36}	2	-2.172	0.2167	-0.930
r^{60}	2	-2.010	0.2823	-0.860
r^{120}	2	-1.791	0.3846	-0.758
1954 : 07 - 1979 : 09				
r^1	2	-1.703	0.4282	-0.312
r^{12}	1	-1.321	0.6202	0.379
r^{36}	1	-1.305	0.6280	0.482
r^{60}	2	-0.887	0.7916	0.939
r^{120}	2	-0.350	0.9141	1.603
1979 : 10 - 2007 : 03				
r^1	14	-3.105	0.0272	-0.639
r^{12}	12	-2.360	0.1539	-0.376
r^{36}	2	-1.472	0.5465	-0.514
r^{60}	2	-1.303	0.6289	-0.462
r^{120}	2	-1.144	0.6987	-0.392

rejection increases with the maturity of the yields. This holds for all yields for the ADF test, while there are two small exceptions with DF-GLS test. The yield with one year to maturity rejects less stronger than the yield with three, five and ten years to maturity for the full sample, as for the second subsample. Moreover, the probability of non-rejection is weaker for the full sample than for both subsamples. While there is also less probability for the period of 1979:10 to 2007:03 than for the period from 1954:07 to 1979:09. The result, that the tests for full sample tend to reject the H_0 more than the first and second subsample may be due to the regime shift. Hence we are in line with Lanne(2000), if we conclude that a structural break is likely for the time series of interest rates. The evidence in favor of a unit root for the first subsample is strong. This finding is in accordance with the theory, which says that interest rates are random walks because of central bank smoothing behaviour[3]. Lanne(2000) sees the lower probabilities of non-rejection in the second subsample due to the small number of observations. We may now reject this assumption, since we have even more data for the second subsample than for the first.

[3] see e.g. Mankiw and Miron (1986)

Correlations It is now worth to have a look on the correlations between the interest rates for the full sample as for the two subsamples. The correlations vary between 0.8466 between the overnight rate and the yield with ten years to maturity of in the first subsample and 0.9961 between the yields with three and five years to maturity also of the first subsample. Especially the correlations containing the overnight rate are relatively low. This may come from the fact that the overnight rate can only be seen as an approximation for a yield with one month to maturity. Nevertheless, also the other correlations are lower than the results Lanne(2000) reports. The correlations are higher for the second subsample than for the first subsample and the full sample. This fact is quite surprising when we compare it with the results from the cointegration tests, which we will see later. Altogether, we have to be aware of the nonrobustness of cointegration methods, which will probably be problematic in this data set.

4.2 Cointegration Tests

I will now give a quick overview about the cointegration tests used. Then a survey about the results is made and some comments on problems of the test procedures are mentioned.

The Johansen Procedure Johansen´s procedure is a multiple equation method and tests on the number of linearly independent cointegration relationships of a set of variables. More precisely it tests for the rank of the cointegration space. We have already seen that in a VECM like in equation (7), this is equal to checking for the rank of π. This is exactly what the Johansen procedure does by applying maximum likelihood to a VECM. The errors are assumed to be Gaussian. The procedure delivers to two test statistics for cointegration. The first one is called trace test and test the hypothesis of there being at most r_0 cointegrating relationships against the hypothesis of more than r_0 relationships. Thus in a set of n variables

$$H_0 : rk(\pi) = r_0 \quad \text{vs.} \quad H_1 : r < rk(\pi) \leq n \quad ,$$

where $r_0 < n$. The second is called maximum eigenvalue statistic and test the null hypothesis of a cointegration rank equal to r_0 against the alternative of a rank equal to $r_0 + 1$, that is

$$H_0 : rk(\pi) = r_0 \quad \text{vs.} \quad H_1 : rk(\pi) = r_0 + 1 \quad .$$

Although, errors are assumed to be Gaussian, the test statistics are asymptotically not χ^2-distributed. The distribution depends on the number of common trends $K - r_0$, as well as on how the procedure is implemented, that is whether an intercept or a deterministic trend are used[4]. In our case no drift is used, and a constant enters only via the error correction term. This procedure was used, since, interest rates do not show any tendency to have a deterministic trend.

LR-test for Restrictions on Cointegration Vectors Johansen also provides a likelihood ratio test for restrictions on cointegration vectors. For simplicity I will call this test, the LR-test. This test also assumes, that errors are Gaussian. For our case it is of interest to restrict the cointegration vectors to be spreads between yields of different maturity. Then we may test whether the spreads span the cointegration space, as it is assumed by our framework. Therefore, the null hypothesis is, that $n-1$ spreads formed by n yields span the cointegration space, all conditional on the rank of cointegration being $n-1$. Under the null hypothesis the test statistic is asymptotically $\chi^2(n-1)$-distributed.

Empirical Results The results for the trace test[5] and the LR-test for restricted cointegration vectors[6] are reported in Table 2 through Table 4. The maximum eigenvalue test was also employed, but the results implied the same rank as the trace tests. Also systems were considered, consisting of only two, three or four yields. The overnight rate was included in every sample, this can be seen as a type of normalization, although tests on the cointegration rank are invariant to normalization and it need not to be reconsidered here. The entire sample displays clear tendency in favor of a rank of $n-1$ for subsets with two or three yields, since the H_0 of a rank equal to $n-1$ can not be rejected, while the H_0 of a rank, that equals $n-2$ can be rejected. But this evidence weakens as the dimension of the system increases. In particular the full system may not reject the H_0 of a rank of only three. For smaller systems the assumption that the spreads between yields span the coitegrating space is rejected on a 5% significance level only for the spread between r_1 and r_{12}. For larger systems the H_0 can not be rejected only for the system $r_1, r_{36}, r_{60}, r_{120}$. We see that, although the λ_{trace}-test can not reject a rank of $n-2$, the LR-test, is not able to reject the H_0, that spreads are the cointegrating vectors. Thus, both test also may be contradictory.

The results of the period from 1954 : 07 to 1979 : 09, presented in Table 3, show a different

[4] Distribution is displayed in the Appendix

[5] critical values for the λ_{trace} and for the $\lambda_{trace}^{Impr.}$ are taken from Osterwald Lenum(1992). For $n-2$ and $n-1$ the 5% critical vaules are 20.26 and 9.16 respectively.

[6] 5% and 10% critical values are 3.84 and 2.71 (two-variable system), 5.99 and 4.61 (three-variable system), 7.81 and 6.25 (four-variable system) and 9.49 and 7.78 (five-variable system), respectively, and degrees of freedom are $n-1$

Table 2: λ_{trace}-test, improved λ_{trace}-test and LR-test for the full sample

Interest rates in the system rank at most	λ_{trace} $n-2$	$n-1$	$\lambda_{trace}^{Impr.}$ $n-2$	$n-1$	LR-test
r_1, r_{12}	31.27	6.30	30.88	6.22	6.55
r_1, r_{36}	29.06	5.67	28.69	5.60	2.18
r_1, r_{60}	28.20	5.37	27.85	5.30	1.40
r_1, r_{120}	26.41	4.57	26.08	4.51	0.69
r_1, r_{12}, r_{36}	32.72	6.53	32.10	6.41	5.13
r_1, r_{12}, r_{60}	28.36	5.95	27.82	5.83	4.45
r_1, r_{12}, r_{60}	25.81	4.74	25.33	4.65	4.59
r_1, r_{36}, r_{60}	20.36	5.08	19.97	4.99	1.61
r_1, r_{36}, r_{120}	20.20	3.70	19.82	3.63	1.77
r_1, r_{60}, r_{120}	21.42	3.31	21.88	3.25	1.34
$r_1, r_{12}, r_{36}, r_{60}$	19.22	5.19	18.74	5.06	9.82
$r_1, r_{12}, r_{36}, r_{120}$	20.12	4.04	19.61	3.94	9.01
$r_1, r_{12}, r_{60}, r_{120}$	22.71	3.91	22.14	3.81	7.31
$r_1, r_{36}, r_{60}, r_{120}$	19.04	3.51	18.56	3.42	2.24
$r_1, r_{12}, r_{36}, r_{60}, r_{120}$	17.77	3.98	17.21	3.86	9.60

pattern. The picture is not quite clear, but there is a tendency that the results evolve in the opposite direction. That is smaller systems may not reject the H_0 of a rank of $n-2$ while, larger dimensioned system do and can not reject the H_0 of $n-1$ linearly independent cointegrating relationships, all on a 5% significance level. However, there is also a tendency that systems containing yields of low maturity can not reject the H_0 of a rank of $n-2$, while systems containing yields with higher maturity can. The LR-test implies for every system, that the spreads span the cointegration space. Also here the λ_{trace}-test and the LR-test are contradictory for some cases. Table 4 shows the results for the second subsample. Here only the systems r_1, r_{12} and r_1, r_{12}, r_{36} can reject the H_0 of a cointegration rank of $n-2$. For increasing dimensions of the systems, as well as for higher maturities, the rejection becomes less possible. The LR-test can not reject the H_0 for any small system, except r_1, r_{12}. Again those results are contradictory to the λ_{trace}-test, this time, however, for almost every small system. For larger systems, also the LR-test rejects the null hypothesis, except for $r_1, r_{36}, r_{60}, r_{120}$. The finding that the cointegration rank is $n-2$ for systems larger than two, would imply that mutual cointegration in those systems is not possible.

The theoretical framework implies for a set of n yields a cointegration rank equal to $n-1$. Results from the Johansen test supporting this implication can only be found for small sys-

Table 3: λ_{trace}-test, improved λ_{trace}-test and LR-test for the first subsample

Interest rates in the system rank at most	λ_{trace} $n-2$	$n-1$	$\lambda_{trace}^{Impr.}$ $n-2$	$n-1$	LR-test
r_1, r_{12}	18.01	3.30	17.53	3.21	3.42
r_1, r_{36}	18.79	3.20	18.29	3.12	2.42
r_1, r_{60}	20.97	3.05	20.41	2.97	1.66
r_1, r_{120}	24.73	3.53	24.08	3.44	0.87
r_1, r_{12}, r_{36}	15.64	3.17	15.03	3.05	3.32
r_1, r_{12}, r_{60}	16.80	2.94	16.14	2.82	2.94
r_1, r_{12}, r_{60}	20.21	3.69	19.41	3.55	2.25
r_1, r_{36}, r_{60}	21.07	2.54	20.23	2.43	3.90
r_1, r_{36}, r_{120}	22.80	5.11	21.90	4.90	5.77
r_1, r_{60}, r_{120}	24.17	7.08	23.21	6.80	5.37
$r_1, r_{12}, r_{36}, r_{60}$	16.15	2.38	15.30	2.26	5.33
$r_1, r_{12}, r_{36}, r_{120}$	23.05	5.27	21.83	4.99	6.08
$r_1, r_{12}, r_{60}, r_{120}$	25.21	5.69	23.88	5.39	5.54
$r_1, r_{36}, r_{60}, r_{120}$	29.19	6.41	27.65	6.07	6.24
$r_1, r_{12}, r_{36}, r_{60}, r_{120}$	23.67	5.68	22.11	5.30	6.95

tems (containing two or three yields) of the full sample and for larger systems (four or five yields) of the first subsample. For the first subsample the λ_{trace}-test rejects this implication in almost every case, while the LR-test implies that the spread vectors are the cointegration vectors. In some cases we found this result, although the λ_{trace}-test does not imply a rank of $n-1$. For the second subsample hardly for any system $n-1$ cointegrating relations can be found, while the LR-test implies that the spread vectors are cointegrating, at least for small systems. For the full sample this outcome only arises for the system $r_1, r_{36}, r_{60}, r_{120}$. Concluding we can state that the λ_{trace}-test and the LR-test are in many cases not in line with theory, although in some cases they are even contradictory. In particular the second subsample does not show evidence in favor of the theoretical implications.

However, the Johansen tests are not very robust. Some characteristics of those tests, will be mentioned now. We have seen, that in some cases the LR-test assumes spreads to be the cointegrating vectors, although the λ_{trace}-test displays a rank of $n-2$ or even lower. This is particularly the case for small systems. Moreover, in the first subsample we found a cointegration rank of $n-1$ for systems, consisting of four or five yields, while for two-variable-systems often no cointegration was found. Podivinsky(1990) showed that all test used may be misleading if too few variables are included. Thus, results of small systems should be

Table 4: λ_{trace}-test, improved λ_{trace}-test and LR-test for the second subsample

Interest rates in the system	λ_{trace}		$\lambda_{trace}^{Impr.}$		LR-test
rank at most	$n-2$	$n-1$	$n-2$	$n-1$	
r_1, r_{12}	21.11	4.25	20.60	4.15	5.15
r_1, r_{36}	19.07	3.94	18.61	3.84	1.12
r_1, r_{60}	17.11	3.89	16.70	3.80	0.85
r_1, r_{120}	14.57	3.53	14.21	3.45	0.52
r_1, r_{12}, r_{36}	20.95	4.30	20.19	4.15	3.00
r_1, r_{12}, r_{60}	15.09	4.12	14.54	3.97	1.25
r_1, r_{12}, r_{60}	12.38	3.63	11.94	3.49	0.60
r_1, r_{36}, r_{60}	10.37	4.02	9.99	3.88	0.07
r_1, r_{36}, r_{120}	9.67	3.84	9.32	3.70	0.32
r_1, r_{60}, r_{120}	10.71	3.48	10.33	3.35	1.58
$r_1, r_{12}, r_{36}, r_{60}$	11.13	4.59	10.60	4.36	15.54
$r_1, r_{12}, r_{36}, r_{120}$	10.57	4.37	10.06	4.16	18.44
$r_1, r_{12}, r_{60}, r_{120}$	11.55	3.67	11.00	3.49	19.72
$r_1, r_{36}, r_{60}, r_{120}$	9.69	3.52	9.22	3.35	6.61
$r_1, r_{12}, r_{36}, r_{60}, r_{120}$	10.07	3.93	9.46	3.69	20.35

regarded cautiously.

For LR-tests the fact that interest rates are not be exact I(1)-processes, at least for the second subsample, may lead to rejection of the H_0, that spreads span the cointegration space, as Lanne(2000) mentions. Our analysis has shown, that on a 5% significance level we can not reject the H_0, that interest rates are I(1)-processes. However, probability in many cases is low. Therefore, it may also be the case, that they are only near a unit root process.

I have already mentioned, that the Johansen tests base on the assumption of error terms to be independently normal. The procedure is very sensitive to this assumption. Rejection of the H_0 and thus finding of a higher cointegration rank is more likely in case of nonnormality. Results from residual analysis show, that in our case we have to deal with the problem of nonnormal errors. Tests on the hypothesis of errors being normally distributed can be rejected for all three samples. Rejection is mainly due to excess kurtosis. Results from Huang and Yang(1996) have shown that with nonnormal errors overestimation of the cointegration rank is more likely for the Johansen procedure than for least square procedures. Therefore, nonnormality of the errors probably can not explain the reason why, specifically for the second subsample, we found a cointegration rank, that is lower, than the one implied by theory.

Both, the λ_{trace} and the λ_{max}-statistic have to deal with a bias. Cheung and Lai(1993) found out that this sample is a positive function of $T/(T - np)$, where T is the sample size, n is the number of yields and p the number of lags used. Therefore Maddala et al.(1998) states that it is appropriate to multiply the test statistics with $(T - np)/T$. I also applied this "improvement". As before a lag length of four was used. The results are reported in the Tables 2 to Tables 4. However, the results do not differ much from those before. They also do not support the theoretical assumptions, since this procedure lowers the test statistics and so non-rejection of a cointegration rank equal to $n - 2$ becomes more probable.

An insufficient lag length may lead to substantial size distortions. Therefore, I also provided the test with a lag length of ten. Results remained the same for the second subsample, while they changed for the first subsample. Now, large system do no more indicate a rank equal to $n - 1$. However, this result may be due to overspecification, which leads to a loss of power, as Boswijk et al.(1992) state.

As said, only the first sample provides evidence in favor of the framework model. However, the second subsample clearly does not. Lanne(2000) refers to the structural break, due to the regime shift, as a plausible factor, why also the entire sample may not be in line with theory. Those structural breaks, however, does not have to be due to regime shifts, also business cycles may drive them. Bansal et al.(2004) mention regime-shifting-models, which can catch up those structural breaks. They also call the time span between 1996 and 2000 a tough challenge for standard term structure models, since it was characterized by several economic recessions and also periods of booms. If they are right, this seems to be a plausible solution, why the framework model fails, at least for the second subsample. Also the entire sample must be influenced, since the second sample covers more than the half of it. We can finally conclude, that the theoretical framework seems to hold only for the first subsample, which is in line with bank-smoothing behavior. But it does not for the second subsample, which might be due to a higher influence of structural breaks like business cycles, after the interest targeting behavior had changed.

5 A Vector Error Correction Model

I will now present an estimation of a VECM for our case of interest rates. Results from cointegration tests induced to apply this estimation only to the first subsample. Some features of a VECM were already outlined in the theoretical framework, but some further points have to be mentioned. It was used a VECM as the one presented in equation (7), and for estimation Johansen´s maximum likelihood was applied. Restrictions are imposed by using selected spreads. The theoretical framework implies that it should not matter which spreads are used. However, for a practical purpose Hall et al.(1992) supposed to use the least correlated ones. This was also done in our case, and the spreads used, thus were $s(12,1,t-1)$, $s(36,12,t-1)$, $s(60,36,t-1)$ and $s(120,60,t-1)$. The test procedure was applied using no determinstic trend and two lagged differences. The results are presented in Table 5. The Johansen LR-test on restricted cointegration vectors shows evidence, that the spreads are the cointegrating vectors, since the H_0 can not be rejected on a 5% significance level[7]. I only

Table 5: Estimation Results from a VECM: 1954:07 to 1979:09

Explanatory Variable	$\Delta r(1,t)$	$\Delta r(12,t)$	$\Delta r(36,t)$	$\Delta r(60,t)$	$\Delta r(120,t)$
$S(12,1,t-1)$	0.061451	-0.034084	-0.024921	-0.029149	-0.016586
t-value	(1.89766)	(-1.22525)	(-1.08753)	(-1.47285)	(-1.09949)
$S(36,12,t-1)$	-0.090005	-0.044490	-0.133718	-0.080028	-0.112222
t-value	(-0.77395)	(-0.44534)	(-1.62487)	(-1.12597)	(-2.07140)
$S(60,36,t-1)$	-0.464122	0.048036	0.327593	0.112230	0.343421
t-value	(-1.30152)	(0.15681)	(1.29820)	(0.51495)	(2.06724)
$S(120,60,t-1)$	0.537289	0.347972	0.168324	0.178487	-0.062963
t-value	(2.99272)	(2.25626)	(1.32492)	(1.626709)	(-0.75281)
	Diagnostic Statistics				
R^2	0.278588	0.256499	0.215691	0.188405	0.189262
S.D. dependent	0.381272	0.322626	0.258759	0.219692	0.167549
Log-likelihood	-86.92598	-41.34425	16.81932	60.79105	142.2354

display the coefficients for spreads, because from an economic point of view long-run behavior is of a higher interest. As mentioned in the frameork part, the coefficients can be seen as adjustment coefficients. They display on how strong the interest rates react on deviations of the equilibrium. Results show that yields react more intensely, if spreads between yields of higher maturity deviate from their equilibrium values, than if spreads between yields of lower maturity do. However, we have to be careful with the results, because many of the coefficients are not statistically significant on a 5% level[8], as can be seen when looking for the t-values printed in parentheses.

6 Conclusion

This paper has showed by means of a theoretical framework, that the idea of term structure of interest rates implies mutual cointegration in a set of n yields of different maturity. Then different test procedures for testing on unit roots and cointegration were introduced. Those procedures then were applied to empirical data. However, the results obtained, many times

[7] $\chi^2(4)$-test statistic is 6.951628 and so non-rejection has a probabilty of 0.138467

[8] T-values can be obtained from any standard t-distribution

were not in line with the theoretical implications, particularly for the second subsample. Only the first subsample showed evidence in favor, which may be due to bank smoothing behavior. Results from the second subsample may reject the implication of our framework, because interest rates are not exact I(1)-processes. For this sample, it is referred to the paper by Bansal et al., which assume unsteady conditions in the nineties as a tough challenge for models like ours. Hence, structural breaks, like business cycles, might be the reason, why results for the second subsample and also for the entire sample are not in line with our theoretical implications. For further studies the application of regime-shifting models is recommended, since they seem to deal better with such structural breaks.

Last but not least a VECM was applied to data from the first subsample. It showed that interest rates tend to react more if spreads between yields of high maturity deviate from there long run equilibria.

In a nutshell, we can conclude that new data has not brought evidence for cointegrating behavior of interest yields, and thus for our theoretical framework. It is rather be the case, that new procedures like regime-shifting models have to be applied for further studies.

7 Appendix

A How to rearrange equation (3):

$$
\begin{aligned}
[r(k,t) - r(1,t)] &= \tfrac{1}{k}\left[\sum_{j=1}^{k} E_t\left(r(1,t+j-1)\right)\right] - r(1,t) + L(k,t) \\
&= \tfrac{1}{k}\Big[E_t\left(r(1,t)\right) + E_t\left(r(1,t+1)\right) + \ldots + E_t\left(r(1,t+k-1)\right) - k \cdot r(1,t)\Big] + L(k,t) \\
&= \tfrac{1}{k}\Big[E_t\left(r(1,t+1) - r(1,t)\right) + E_t\left(r(1,t+2) - r(1,t)\right) + \ldots \\
&\quad + E_t\left(r(1,t+k-1) - r(1,t)\right)\Big] + L(k,t) \\
&= \tfrac{1}{k}\left[\sum_{j=1}^{k-1} E_t\left[r_(1,t+j) - r(1,t)\right]\right] + L(k,t) \\
&= \tfrac{1}{k}\Big[\sum_{j=1}^{k-1} E_t\left[r_(1,t+1) - r(1,t)\right] + E_t\left[r_(1,t+2) - r(1,t+1)\right] + \ldots \\
&\quad + E_t\left[r(1,t+i) - r(1,t+i-1)\right]\Big] + L(k,t) \\
&= \tfrac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{j=i} E_t \Delta r(1,t+j) + L(k,t)
\end{aligned}
$$

B Test distribution of the ADF test: The asymptotic distribution of $T\hat{\rho}$ and $T\hat{\theta}$ are independent. This result can be used to show, that the distribution of $\hat{\rho}$ is the Dickey-Fuller distribution, while the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta)$ is normal. The t-statistic for $\hat{\rho}$ then is

$$t_{\hat{\rho}} \Rightarrow \frac{\int W{*}dW}{\int W{*}^2 dr} \quad ,$$

where $W{*} = W(r) - \int W(r)dr$ is the "demeaned" Brownian Motion.

C Distribution of the λ_{trace}-statistic:

$$\int (dW)W{*}' \left(\int W{*}W{*}'\right)^{-1} \int W{*}(dW)' \quad ,$$

where $W{*}$ is again the "demeaned" Brownian Motion.

References

[1] Bansal, R., Tauchen, G. and Zhou, H. (2004). "Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle". *Journal of Business & Economic Statistics*, 22, 396-408.

[2] Boswijk, H.P. (1992). *Cointegration, Identification and Exogeneity.* Amsterdam: Tinbergen Institute Thesis Publishers.

[3] Cheung, Y., and Lai, K. (1993). "Finite Sample Sizes of Johansen´s Likelihood Ratio Tests for Cointegration". *Oxford Bulletin of Economics and Statistics*, 55, 313-328.

[4] Choi, I., and Chung, B. (1995). "Sampling Frequency and the Power of Tests for a Unit Root: A Simulation Study". *Economics Letters*, 49, 131-136.

[5] Eitrheim, O. (1992). "Inference in Small Cointegrated Systems: Some Monte Carlo Results". Presented at the Economic Society Meeting in Brussels.

[6] Elliott, G., Rothenberg, T.J., and Stock, J.H. (1996) "Efficient Tests for an Autoregressive Unit Root". *Econometrica*, 64, 813-836.

[7] Engle, R.F., and Granger C.W.J. (1987). "Co-integration, and Error Correction: Representation Estimation and Testing". *Econometrica*, 66, 149-158.

[8] Hall, A.D., Anderson, H.M., and Granger, C.W.J. (1992). "A Cointegration Analysis of Treasury Bill Yields". *Review of Economics and Statistics*, 74, 116-126.

[9] Hamilton, J.D. (1994). *Time Series Analysis.* New Jersey: Princeton University Press.

[10] Huang, B.N., and Yang, C.W. (1996). "Long-run Purchasing Power Parity Revisited: A Monte Carlo Simulation" *Applied Economics*, 28, 967-974.

[11] Lanne, M.(2000). "Near Unit Roots, Cointegration, and the Term Structure of Interest Rates". *Journal of Applied Econometrics*, 15, 513-529.

[12] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Berlin: Springer Presse.

[13] Lütkepohl, H., and Krätzig, M. (2004). *Applied Time Series Econometrics.* Cambridge: Cambridge University Press.

[14] Maddala, G., and Kim, I.M. (1998) *Unit roots, cointegration and structural change.* Cambridge : Cambridge University Press.

[15] Mankiw, N.G., and Miron, J.A. (1986). "The Changing Behavior of the Term Structure of Interest Rates". *Quarterly Journal of Economics*, 101, 211-228.

[16] Osterwald-Lenum, M. (1992). "A Note with fractiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics: Four Cases". *Oxford Bulletin of Economics and Statistics*, 54, 461-72.

[17] Podivinsky, J.M. (1990). "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation". *Econometric Theory*, 8, 1-27.

[18] Stock, J.S. and Watson, M. (1988). "Variable Trends in Economic Time Series". *Journal of Economic Perspectives*, 2.