Unstable Multiple Cointegration Relations in the Term Structure of Interest Rates

Heejoon Kang [1]

Although the literature has theoretically shown that multiple cointegration relations are not uniquely defined, many empirical analyses report and make use of such multiple cointegrations. This paper shows that four long- maturity interest rates in the United States contain two common factors and cointegration rank is thus two. Multiple cointegration relations among four interest rates are unstable and sensitive to small changes in the number of observations. Through Monte Carlo sampling experiments, the nature and the extent of instability are established. Instead of multiple cointegration relations, stable irreducible cointegration relations among three interest rates are presented.

Keywords: Cointegration rank, Irreducible cointegration, Maximum likelihood estimation

JEL Classification: C32, C40, E43

I. Introduction

Multiple cointegrations occur when there is more than one cointegration relation among more than two nonstationary, I(1), time series. For instance, among four I(1) time series, there are four possibilities: no cointegration, one cointegration relation, two cointegration relations, or three cointegration relations. That is, the rank of cointegration is either 0, 1, 2, or 3. The literature is abundant with discussions for zero or one cointegration relation.

*Professor of Business Economics and Public Policy, Kelley School of Business, Indiana University, Bloomington, IN 47405, USA, (Tel) +1-812-855-9219, (E-mail) kang@indiana.edu.
[**Seoul Journal of Economics** 2002, Vol. 15, No. 1]

This paper will investigate the issues of identification and estimation when there are multiple cointegrations such that the cointegration rank is either two or three among four I(1) series. In particular, it will be shown that cointegration relations in multiple cointegrations are unstable: estimated parameter values are sensitive to small changes in the number of observations. The situation is similar to the case of multicollinearity in multiple regressions. A drop of one or two observations, say, out of 100 observations, would change the values of estimated regression parameters substantially. In other words, the estimated cointegration relations are unstable and unreliable.

It should be noted here that multiple cointegrations discussed in this paper are different from multicointegration in Granger and Lee (1990), who describe the multiple layers of cointegration. Some I(2) time series, for instance, may be cointegrated so that a linear combination becomes I(1). This particular linear combination of I(1) can be further cointegrated with some other I(1) time series in the multicointegration.

Multiple cointegrations may occur in a simultaneous equations model (SEM). Four I(1) endogenous series may be determined by a four-equations SEM, in which some exogenous variables are I(1) and some are I(0). Certain combinations of the four endogenous variables are I(0), and some other combinations are I(1). When a SEM is fully specified, the fact that certain combinations are I(1) or I(0) is not essential, but only a by-product of the model. That is, a SEM would show much fuller relations among endogenous and exogenous variables than a mere cointegration investigation. Naturally, we assume that such a SEM is not provided. Among four I(1) series, the inquiry in this paper is whether some combinations are either I(1) or I(0). Many analysts have reported the presence of multiple cointegrations in various applications. The rank of cointegration is typically investigated through the procedures developed by Johansen (1988, 1991), Ahn and Reinsel (1990), and Johansen and Juselius (1990).

The objective of this paper is to identify and estimate cointegration relations in the term structure of United States interest rates. It will be shown that multiple cointegration relations are not identified. That is, there is no unique way to characterize a cointegration relation when there is more than one such relation. The significance of multiple cointegrations is shown to be the existence of cointegration among a subset of time series. That is, when there are two cointegration relations among four I(1) series, then any three of those four series will in general be also cointegrated. Moreover, no unique cointegration relations can be estimated by means of the conventional maximum likelihood estimation (MLE) method when there are multiple cointegrations. Instead of multiple cointegration relations, irreducible cointegration relations, suggested by Wickens (1996), for a minimum set of cointegrated I(1) time series should be investigated. We will make use of the concept of latent common factors to establish the nature of multiple cointegrations. Common I(1) factors are nonstationary time series that would generate the given set of cointegrated time series.

The structure of the paper is as follows. Section II formally describes the role of latent common factors in multiple cointegrations. Section III adopts the concept of irreducible cointegration by Wickens (1996) to show how multiple cointegration relations imply cointegration relations among certain subsets of time series. Section IV applies multiple cointegrations and irreducible cointegrations to the United States term structure of interest rates. Monthly long-maturity (five-year, seven-year, ten-year, and 30-year) rates from 1982:01 to 1999:12 are shown to contain two I(1) common factors. Therefore, the cointegration rank is two and stable irreducible cointegrations exist among every possible set of three long-maturity rates. After illustrating how sensitive estimated multiple cointegration relations are to the number of observations in Section IV, a series of Monte Carlo sampling experiments are conducted to show the nature and the extent of parameter instability in Section V. Concluding remarks will be given in Section VI.

II. Multiple Cointegrations and Latent Common Factors

Consider a relationship, following Engle and Granger (1987), between the two I(1) time series of y_{1t} and y_{2t}:

$$y_{1t}=z_t+\varepsilon_{1t}, \qquad y_{2t}=\beta z_t+\varepsilon_{2t}, \tag{1}$$

where z_t is I(1), and ε_{1t} and ε_{2t} are I(0). Both y_{1t} and y_{2t} are I(1) due to the dominance of z_t over ε_{1t} and ε_{2t}. In general, a linear combination of $y_{2t}-\lambda y_{1t}$ becomes $(\beta-\lambda)z_t+\varepsilon_{2t}-\lambda\varepsilon_{1t}$, which is also I(1). A particular combination of $y_{2t}-\beta y_{1t}$, however, becomes $\varepsilon_{2t}-\beta\varepsilon_{1t}$, which is I(0) by construction. Hence, y_{1t} and y_{2t} in (1) are cointegrated.

If z_t in (1) were known or observed, then the fact that y_{1t} and y_{2t} are cointegrated would be trivial, because it should be known how y_{1t} is determined by and how y_{2t} is determined by. A linear combination between the two series being cointegrated would be a by-product in the sense that both series happen to share a common I(1) determinant. It is important to note that the cointegration relation of $y_{2t}-\beta y_{1t}$ is still I(0) *even if* z_t is *not observed* and hence it is *latent*. Under the assumption that z_t in (1) is not observed, it will be called a latent common factor, because this latent factor commonly determines both y_{1t} and y_{2t}. Between the two I(1) series, they are either cointegrated or not cointegrated. In order to meaningfully discuss multiple cointegrations, several situations among four I(1) series will be discussed.

If there is one cointegration relation among four I(1) series, the system will in general have three common factors as in

$$\begin{aligned} y_{1t}&=\beta_1 z_{1t}+\beta_2 z_{2t}+\beta_3 z_{3t}+\varepsilon_{1t},\\ y_{2t}&=z_{1t}+\varepsilon_{2t},\\ y_{3t}&=z_{2t}+\varepsilon_{3t},\\ y_{4t}&=z_{3t}+\varepsilon_{4t}, \end{aligned} \tag{2}$$

z_t's are I(1) (which are not themselves cointegrated), and ε_{it}'s are I(0).

The coefficients of z_t's for y_{2t}, y_{3t}, and y_{4t} are normalized to unity. As long as z_t's are not observable, (2) is equivalent to, for instance,

$$\begin{aligned} y_{1t}&=\beta_1 z_{1t}+\beta_2 z_{2t}+\beta_3 z_{3t}+\varepsilon'_{1t},\\ y_{2t}&=\beta_1' z_{1t}+\beta_2' z_{2t}+\beta_3' z_{3t}+\varepsilon'_{2t},\\ y_{3t}&=\beta_1'' z_{1t}+\beta_2'' z_{2t}+\beta_3'' z_{3t}+\varepsilon'_{3t},\\ y_{4t}&=\beta_1''' z_{1t}+\beta_2''' z_{2t}+\beta_3''' z_{3t}+\varepsilon'_{4t}. \end{aligned} \tag{3}$$

Without loss of generality, (3) can be written as (2) and therefore there is one cointegration relation, namely, $y_{1t}-\lambda_1 y_{2t}-\lambda_2 y_{3t}-\lambda_3 y_{4t}$.

When the rank of cointegration is two so that there are two cointegration relations, then the system has two latent common factors of z_{1t} and z_{2t}:

$$\begin{aligned} y_{1t} &= \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon'_{1t}, \\ y_{2t} &= \beta_1' z_{1t} + \beta_2' z_{2t} + \varepsilon'_{2t}, \\ y_{3t} &= \beta_1'' z_{1t} + \beta_2'' z_{2t} + \varepsilon'_{3t}, \\ y_{4t} &= \beta_1''' z_{1t} + \beta_2''' z_{2t} + \varepsilon'_{4t}. \end{aligned} \tag{4}$$

which can also be written as, without loss of generality,

$$\begin{aligned} y_{1t} &= \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon_{1t}, \\ y_{2t} &= \beta_1' z_{1t} + \beta_2' z_{2t} + \varepsilon_{2t}, \\ y_{3t} &= z_{1t} + \varepsilon_{3t}, \\ y_{4t} &= z_{2t} + \varepsilon_{4t}, \end{aligned} \tag{5}$$

It is important to note, however, that any three time series of y_t's in (5) will also be cointegrated. Any three time series will share two common factors so that there will be a unique cointegration relation. A cointegration relation among four time series is, however, not unique. In fact, there is an infinite number of cointegrating relations among the four variables. Wickens (1996), Rossana (1998), Luukkonen, Ripatti, and Saikkonen (1999), and others state that cointegration relations are in general not identified when there are more than one relation.

When the rank of cointegration is three so that there are three cointegration relations, then there will be only one latent common factor. Without loss of generality such a system can be written as

$$\begin{aligned} y_{1t} &= \beta_1 z_t + \varepsilon_{1t}, \\ y_{2t} &= \beta_2 z_t + \varepsilon_{2t}, \\ y_{3t} &= \beta_3 z_t + \varepsilon_{3t}, \\ y_{4t} &= z_t + \varepsilon_{4t}. \end{aligned} \tag{6}$$

In (6), any two time series will be cointegrated. Cointegration relations between any two time series are well defined, but those among either three or four time series are not identified. That is, there will be infinitely many solutions to make $y_{1t} - \lambda_1 y_{2t} - \lambda_2 y_{3t} - \lambda_3 y_{4t}$ into an I(0) series.

Finally, a system of four I(1) time series, which are not cointegrated, is

$$
\begin{aligned}
y_{1t} &= z_{1t} + \varepsilon_{1t}, \\
y_{2t} &= z_{2t} + \varepsilon_{2t}, \\
y_{3t} &= z_{3t} + \varepsilon_{3t}, \\
y_{4t} &= z_{4t} + \varepsilon_{4t},
\end{aligned}
\tag{7}
$$

where there are no common factors. Since the only requirement is that z_t's should be I(1) (which are not themselves cointegrated) and ε_t's are I(0), (7) can be further reduced to

$$
\begin{aligned}
y_{1t} &= z_{1t}', \\
y_{2t} &= z_{2t}', \\
y_{3t} &= z_{3t}', \\
y_{4t} &= z_{4t}',
\end{aligned}
\tag{8}
$$

by absorbing ε_t's into z_t's.

III. Irreducible Cointegrations

Wickens (1996) introduces the term, "irreducible cointegration," to denote a set of minimum number of I(1) time series, which are cointegrated. For instance, if the rank of cointegration is two among four I(1) series, then in general any three I(1) series will be cointegrated. Any two I(1) series, however, will not be cointegrated so that irreducible cointegration occurs with three series. If the rank of cointegration is three among four I(1) series, then irreducible cointegration occurs with any two time series. Of course, if the rank of cointegration is one, then the original system of four time series will also constitute an irreducible cointegration. The concept of irreducible cointegration is important to identify cointegration relations. If cointegration is not irreducible, then there are multiple cointegrations, of which relations are not identified. For instance, if three time series are of an irreducible cointegration, then an addition of any I(1) series will also be cointegrated, because the weight for the additional series could take a value of zero. For irreducible cointegrations, cointegration relation is uniquely defined, within a multiplicative factor. That is, if $y_{1t} - \beta y_{2t}$

is a cointegration relation so that $[1, -\beta]$ is the cointegration vector, then $[\lambda, -\lambda\beta]$ will also be a cointegration vector with a non-zero λ. Yet, if cointegration is not irreducible to have more than one cointegration relation, then there will be infinitely many vectors, beyond the multiplicative factor. This is because a linear combination of two or more I(0) cointegration relations is in general also an I(0) series. The fact that multiple cointegration vectors are not uniquely determined is well established in the literature. Nevertheless, in applied cointegration analysis, many analysts report and make use of multiple cointegrations, although they are not uniquely identified and therefore arbitrary.

Since the cointegration relations are unique within the multiplicative factor, a particular normalization is needed. For instance, suppose $\beta_1 y_{1t} + \beta_2 y_{2t} + \beta_3 y_{3t}$ is I(0) so that $[\beta_1, \beta_2, \beta_3]$ is a cointegration vector. Normalization is made to have, say, $\beta_1 = 1$ so that a unique vector is obtained as $[1, \beta_1', \beta_2']$. The situation is different under the multiple cointegrations. If the rank of cointegration is two for three I(1) series, then there are infinitely many cointegration vectors. Normalization alone will not obtain a unique vector. That is, even with normalization, $[1, \beta_1, \beta_2]$, there are still infinitely many combinations of β_1 and β_2 to satisfy the cointegration condition. In such cases, a further normalization such as $[1, 1, \beta]$ will work to obtain a unique vector. Under irreducible cointegrations, one of the last two values is further restricted to be zero, for instance, to have $[1, \beta, 0]$ by removing a third time series.

IV. Term Structure of Interest Rates in the United States

A. The data

Monthly interest rates are retrieved from St. Louis Federal Reserve Bank database, FRED, from January 1982 to December 1999 for long-maturity rates of five-year, seven-year, ten-year, and 30-year rates. Below all the interest rates are transformed through logarithmic transformations of log (1.+annualized rate/100). Those four interest rates, after the transformations, are denoted as $Y5$, $Y7$, $Y10$, and $Y30$. As typically done in the literature, the presence of unit roots is investigated by means of the Dickey-Fuller test procedure. Each of four rates has one unit root at the 5% significant level. According to the literature, interest rates are sometimes I(0) and sometimes I(1), depending on the maturity and depending on the time period. For the sample period studied, those four long-maturity interest rates are all I(1).

In the literature, many researchers in fact report that various configurations of interest rates contain multiple cointegrations. For the system of k interest rates, the order of cointegration is reported to be k-1. Stock and Watson (1988) show that the federal funds rate, 90-day Treasury bill rate, and one-year Treasury bill rate in the United States have one common I(1) factor. Engsted and Tanggaard (1994) show that eleven interest rates, ranging from one-month to 15-year rates for Danish bond market, contain one common I(1) factor. Every pair of eleven interest rates is cointegrated. Cuthbertson, Hayes, and Nitzsche (1998) report that one-, three-, six-, and 12-month interest rates in the United Kingdom have one common I(1) factor and that one-, two-, three-, six-, and 12-month rates in Germany also have only one common I(1) factor. For long-maturity interest rates, Bowe and Mylonidis (1999) show that three-, five-, seven-, and ten-year European ECU interest rates share one common I(1) factor. These researchers all report that the (irreducible) cointegration is found between any two interest rates.

The literature frequently reports, only implicitly though, that multiple cointegration relations are unstable. Bradley and Lumpkin (1992) investigate the cointegration relations among seven United States Treasury rates from three-month to 30-year rates. They estimate cointegration relations by regressing each of seven rates as the dependent variable on the other six rates. When those seven regression relations obtained from ordinary least squares (OLS) method are re-expressed such that the same rate appears on the left-hand side, the regression coefficients are very different. That is, the cointegration relations from OLS are very sensitive to the normalization. There are many analysts listed above who report that the number of common factors is greater than one in the term structure of interest rates. That is, irreducible cointegration would contain more than two interest rates. The nature of irreducible cointegration depends on the maturity and the time period of investigation.

In general, the rank of cointegration is found to be greater than one when more than two interest rates of different maturity are analyzed as in Arshanapalli and Doukas (1994), Choi and Wohar (1991, 1995), Engsted (1993), Guest and McLean (1998), Hiraki, Shiraishi, and Takezawa (1996), Mandeno and Giles (1995), McDermott (1998), McFadyen, Pickerill, and Devaney (1991), Shea (1992), Wallace and Warner (1993), and Warne (1997) for various maturities and for many countries. That is, multiple cointegrations have been commonly reported for the term structure of interest rates.

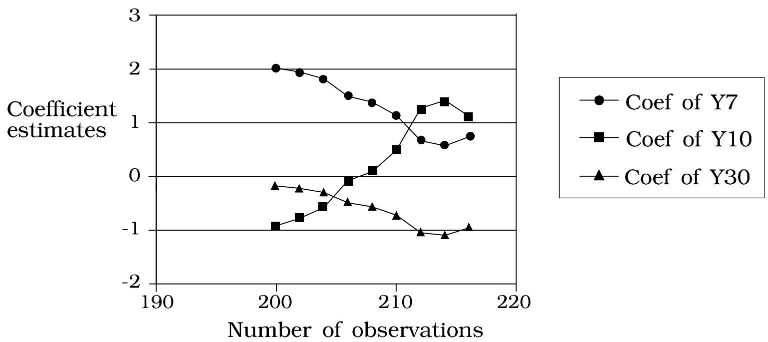
The maximum likelihood estimation (MLE) method by Johansen and Juselius (1990) reveals that the order of cointegration is two among the four rates of *Y*5, *Y*7, *Y*10, and *Y*30. There are two common I(1) factors for these four rates. Any three out of four long-maturity rates are therefore cointegrated. Needless to say, the system of all four rates is also cointegrated. All the analyses for the interest rates are conducted by using RATS (Regression Analysis of Time Series), Version 4, and CATS (Cointegration Analysis of Time Series). The cointegration results from the MLE, by using the lag length of one throughout, are provided in Table 1 for various sample periods. The first one uses all 216 observations from 1982:01 to 1999:12. Subsequently, the number of observations is reduced by two by eliminating the most recent two months from the sample period. The number of observation runs from 216 at the top to 200 at the bottom.

In Table 1, cointegration relations are all expressed such that the same five-year rate (*Y*5) is written in terms of the other rates. The MLE results from various numbers of observations show that the cointegration relations are indeed very sensitive to the number of observations. For instance, the coefficient of the seven-year rate (*Y*7) changes from 0.57751 with 214 observations to 1.95858 with 200 observations. The latter coefficient is almost 3.4 times the former value, although the number of observation differs by 14 observations, which is about 6.5%. The coefficient of ten-year rate (*Y*10) shows even larger variations from -0.84892 with 200 observations to +1.37134 with 214 observations. Similarly, the coefficient of 30-year rate (*Y*30) varies from -1.01217 with 214 observations to -0.12963 with 200 observations. The former is about 7.8 times greater than the latter in absolute values. Three sets of coefficients are plotted in Figure 1. The number of observations in Figure 1 increases from 200 to 216. The picture

TABLE 1

MULTIPLE COINTEGRATIONS AMONG FOUR LONG-MATURITY INTEREST RATES.

$Y5$ = five-year rate; $Y7$ = seven-year rate; $Y10$ = ten-year rate; and $Y30$ = 30-year rate
Between 1982:01 and 1999:12 (216 observations)
$Y5 = \text{constant} + 0.73579\ Y7 + 1.09165\ Y10 - 0.88241\ Y30$
Between 1982:01 and 1999:10 (214 observations)
$Y5 = \text{constant} + 0.57751\ Y7 + 1.37134\ Y10 - 1.01217\ Y30$
Between 1982:01 and 1999:08 (212 observations)
$Y5 = \text{constant} + 0.65880\ Y7 + 1.24120\ Y10 - 0.96013\ Y30$
Between 1982:01 and 1999:06 (210 observations)
$Y5 = \text{constant} + 1.10893\ Y7 + 0.50001\ Y10 - 0.65215\ Y30$
Between 1982:01 and 1999:04 (208 observations)
$Y5 = \text{constant} + 1.34852\ Y7 + 0.12215\ Y10 - 0.50659\ Y30$
Between 1982:01 and 1999:02 (206 observations)
$Y5 = \text{constant} + 1.45514\ Y7 - 0.04830\ Y10 - 0.43950\ Y30$
Between 1982:01 and 1998:12 (204 observations)
$Y5 = \text{constant} + 1.75765\ Y7 - 0.52728\ Y10 - 0.25337\ Y30$
Between 1982:01 and 1998:10 (202 observations)
$Y5 = \text{constant} + 1.88273\ Y7 - 0.72039\ Y10 - 0.18404\ Y30$
Between 1982:01 and 1998:08 (200 observations)
$Y5 = \text{constant} + 1.95858\ Y7 - 0.84892\ Y10 - 0.12963\ Y30$



Model: Y5 = constant + (Coef of Y7)*Y7 + (Coef of Y10)*Y10 + (Coef of Y30)*Y30

FIGURE 1

UNSTABLE COINTEGRATION RELATIONS AMONG FOUR LONG-MATURITY RATES

Note: The estimates are from Table 1 and the number of observation changes from 216 to 200 (but, reversed in the figure).

TABLE 2
IRREDUCIBLE COINTEGRATIONS AMONG FOUR LONG-MATURITY INTEREST RATES

Among $Y5$ (five-year rate), $Y7$ (seven-year rate), and $Y10$ (ten-year rate)
216 observations: $Y5 = \text{constant} + 1.82453\ Y7 - 0.82845\ Y10$
210 observations: $Y5 = \text{constant} + 1.93856\ Y7 - 0.94373\ Y10$
205 observations: $Y5 = \text{constant} + 2.08138\ Y7 - 1.08897\ Y10$
200 observations: $Y5 = \text{constant} + 2.14816\ Y7 - 1.16141\ Y10$
Among $Y5$ (five-year rate), $Y7$ (seven-year rate), and $Y30$ (30-year rate)
216 observations: $Y5 = \text{constant} + 1.36815\ Y7 - 0.39567\ Y30$
210 observations: $Y5 = \text{constant} + 1.40629\ Y7 - 0.43789\ Y30$
205 observations: $Y5 = \text{constant} + 1.42895\ Y7 - 0.46382\ Y30$
200 observations: $Y5 = \text{constant} + 1.43122\ Y7 - 0.46771\ Y30$
Among $Y5$ (five-year rate), $Y10$ (ten-year rate), and $Y30$ (30-year rate)
216 observations: $Y5 = \text{constant} + 2.35824\ Y10 - 1.44853\ Y30$
210 observations: $Y5 = \text{constant} + 2.37656\ Y10 - 1.46891\ Y30$
205 observations: $Y5 = \text{constant} + 2.36611\ Y10 - 1.45972\ Y30$
200 observations: $Y5 = \text{constant} + 2.33581\ Y10 - 1.42045\ Y30$
Among $Y7$ (seven-year rate), $Y10$ (ten-year rate), and $Y30$ (30-year rate)
216 observations: $Y7 = \text{constant} + 1.75700\ Y10 - 0.80527\ Y30$
210 observations: $Y7 = \text{constant} + 1.72931\ Y10 - 0.77519\ Y30$
205 observations: $Y7 = \text{constant} + 1.66958\ Y10 - 0.71260\ Y30$
200 observations: $Y7 = \text{constant} + 1.61990\ Y10 - 0.65234\ Y30$

shows that all three coefficients vary substantially with small changes in the number of observations. That is, cointegration relations in multiple cointegrations are unstable and unreliable. The nature of instability in multiple cointegrations is very similar to that of multicollinearity in multiple regressions (Greene 2000, p. 256).

On the other hand, Table 2 shows the irreducible cointegrations for long-maturity interest rates. Since there are two common I(1) factors among the four interest rates, or in turn, the cointegration rank is two, any three of the four interest rates are cointegrated. Table 2 shows the MLE (with lag length one) results for the four different sets of irreducible cointegrations. There are four subsets of three interest rates, for which the MLE results are provided in each panel. In each panel, four different numbers of observations are tried. The first one is by using all 216 observations, the next one is from 210 by eliminating the most recent six observations. The next two MLE results are, respectively, from 205 and 200 observations, by again eliminating the appropriate number of the most recent observations from the investigation.

The irreducible cointegration relations are stable in Table 2. They are very stable especially when they are compared with those variations in Table 1. The variation of the coefficients in the cointegration relations is rather small. Among five-year rate ($Y5$), ten-year rate ($Y10$), and 30-year rate ($Y30$) given under the third Panel, both the variations from the smallest to the largest coefficients of $Y10$ and those of $Y30$ are less than 2% when the number of observations change from 216 to 200. The coefficient of the ten-year rate ($Y10$) in the first Panel shows the largest variation: the largest coefficient from 200 observations is about 40% bigger than the smallest coefficient, both in absolute values, from 216 observations. Though the differences among the coefficients are not tested for their statistical differences, the largest change of 40% is much smaller than those tremendous changes in Table 1. Table 2 shows that the irreducible cointegration relations are much more stable than multiple cointegrations.

In general, cointegration relations will also depend on the nature of cointegrated time series. It is possible that there might be structural changes within the sample period to exhibit (large) fluctuations in the estimation. In order to firmly establish the nature and the extent of the coefficient stability in both multiple cointegrations and irreducible cointegrations, a series of sampling experiments will be conduced next.

V. Monte Carlo Experiments

Though strongly suggestive, the results in Tables 1 and 2 do not concretely show whether the instability is in fact due to the presence of the multiple cointegrations. Even if there is a unique cointegration, cointegration relations can be unstable if there is, for instance, a structural change in the sample. In order to show the nature and the extent of the parameter instability, a series of Monte Carlo experiments are conducted. In the first series of experiments, three I(1) time series are generated such that there are two cointegration relations. Multiple cointegration relations are estimated by the MLE with different numbers of observations. In particular, the following model is used for the multiple cointegrations:

$$\begin{aligned} y_{1t} &= \beta_1 z_t + \varepsilon_{1t}, \\ y_{2t} &= \beta_2 z_t + \varepsilon_{2t}, \\ y_{3t} &= z_t + \varepsilon_{3t}. \end{aligned} \tag{9}$$

In order to make the data generating process (DGP) simple, z_t is generated as a random walk series and each of ε_t's is independently and identically distributed (iid) normal deviates with the standard deviations of 5.0. Moreover, all three ε_t's are separately generated to make them mutually uncorrelated. By using FORTRAN, three series of 100 observations are generated from the standard normal distribution by using an IMSL (International Mathematical and Statistical Library) subroutine, rnnor. The common factor of z_t is then generated by

$$z_t = 1.0 + z_{t-1} + \varepsilon_{4t}, \tag{10}$$

where ε_{4t} is an additional iid normal deviate with the standard deviation of 10.0. The initial value of zero is used for z_t so that it becomes a random walk series.

Once three series of y's are generated from (9), the presence of a unit root is checked by using the Dickey-Fuller test procedure. If any of three y_t's does not contain a unit root, then the particular replication is ignored from further analyses. This step guarantees that the time series are indeed I(1). Once those three time series are found to be I(1), then they are used to estimate one cointegration relation. Although there are two cointegration relations, the MLE procedure of Johansen and Juselius (1990), which is detailed in Hamilton (1994, pp. 635-8) for step by step instructions, would produce one single cointegration relation. Once the cointegration relation is found by the MLE with the lag length of one, the residuals are checked to see if indeed the given linear combination is I(0). The entire process is repeated 10,000 times in the Monte Carlo experiment.

The estimated results for four different numbers of observations − 100, 98, 96, and 94 − are listed in Table 3. When the certain number of observations is eliminated, the last several observations

TABLE 3

UNSTABLE COINTEGRATION RELATIONS IN MULTIPLE COINTEGRATIONS

True Model: $y_{1t}=1.5\ z_t+\varepsilon_{1t}$; $y_{2t}=0.8z_t+\varepsilon_{2t}$; and $y_{3t}=z_t+\varepsilon_{3t}$.
Estimation: $y_{1t}-a-by_{2t}-cy_{3t}=\mathrm{I}(0)$. The values of a, b, and c are not well defined.

Panel *A*: MLE results

	(1)	(2)	(3)	(4)
NOB	100	98	96	94
a	1.3782 (113.23)	-3.4942 (293.23)	-0.5251 (57.17)	-0.6387 (58.16)
b	0.6242 (78.66)	1.8604 (619.82)	0.1279 (104.70)	1.7104 (103.78)
c	0.9908 (63.21)	0.1323 (488.01)	1.4257 (86.46)	0.1383 (82.33)

Panel *B*: Correlation coefficients

a(NOB)	a(100)	a(98)	a(96)	a(94)
a(100)	1.00000	0.00404 [0.6862]	0.02303 [0.0213]	-0.00674 [0.5002]
a(98)		1.00000	0.03349 [0.0008]	0.00702 [0.4826]
a(96)			1.00000	0.00586 [0.5578]
b(NOB)	b(100)	b(98)	b(96)	b(94)
b(100)	1.00000	0.00637 [0.5240]	0.02723 [0.0065]	0.01255 [0.2094]
b(98)		1.00000	0.00637 [0.5010]	-0.00776 [0.4376]
b(96)			1.00000	-0.17813 [0.0001]
c(NOB)	c(100)	c(98)	c(96)	c(94)
c(100)	1.00000	0.00617 [0.5370]	0.02604 [0.0092]	0.01258 [0.2085]
c(98)		1.00000	0.00649 [0.5163]	-0.00792 [0.4283]
c(96)			1.00000	-0.17098 [0.0001]

Note: Values in parentheses are standard deviations over 10,000 replications and those in brackets are p-values to test if correlation coefficients are zero.

are taken out of the investigation. That is, when the number of observation is 96, the sample consists of the first 96 observations out of the total 100 observations originally generated. Regardless of the number of observations, each of 10,000 replications is found to be indeed cointegrated.

It should be noted that the multiple cointegration relations in Table 3 are not well defined and therefore no specific values are to be expected for the coefficients. The values in the table are the mean values of a, b, and c over those 10,000 replications in the experiment. Values in parentheses are standard deviations over the same 10,000 replications in Panel *A*. For any coefficient, the mean values are very different from one column to another. The drop of two, four, or six observations from the original 100 makes a large difference in the mean values for the cointegration relations. Equally important is the large values of the standard deviations. Even with the same number of observations, multiple cointegration relations vary tremendously from one replication to another.

In Panel *B* of Table 3, simple Pearson correlation coefficients are computed for each coefficient for the given number of observations in order to investigate the stability of the coefficients. For instance, for the coefficient of c, the correlation coefficient between c from 100 observations and c from 98 observations is 0.00617, which is not significantly different from zero. The statistical significance of correlation coefficients is given in brackets for their p-values. Some correlation coefficients are statistically different from zero. For instance, b from 100 observations and b from 96 observations have the correlation coefficient of 0.02723, which has the p-value of 0.0065 indicating that it is significant at the 1% significance level. But the coefficient is not numerically very different from zero. Some correlation coefficients are even negative. They are all numerically close to zero indicating the extreme instability of cointegration relations in this experiment of the multiple cointegrations.

Different measures of coefficient stability are computed. In Table 4, mean deviations, mean absolute deviations, and mean squared deviations are computed over the same 10,000 replications. Three columns are for the estimates of a, b, and c. The first row in each panel is for the difference between the results from 100 observations and those from 98 observations. For instance, the mean absolute deviation for b between 100 observations and 96 observations is 9.4415. Again, this value is very large indicating that the

TABLE 4

DIFFERENCES IN THE PARAMETER ESTIMATES IN MULTIPLE COINTEGRATIONS

True Model: $y_{1t}=1.5z_t+\varepsilon_{1t}$; $y_{2t}=0.8z_t+\varepsilon_{2t}$; and $y_{3t}=z_t+\varepsilon_{3t}$.

Estimation: $y_{1t}-a-by_{2t}-cy_{3t}=\mathrm{I}(0)$.

Panel *A*: Mean deviations			
	a	*b*	*c*
Mean deviations (x100-x98)	4.8724	-1.2361	0.8586
Mean deviations (x100-x96)	1.9033	0.4963	-0.4349
Mean deviations (x100-x94)	2.0169	-1.0861	0.8526
Panel *B*: Mean absolute deviations (MAD)			
	a	*b*	*c*
MAD (x100-x98)	12.3985	17.3808	13.8260
MAD (x100-x96)	8.4817	9.4415	7.6024
MAD (x100-x94)	8.2247	8.9030	7.1254
Panel C: Mean squared deviations (MSD)			
	a	*b*	*c*
MSD (x100-x98)	98547.84	389705.48	241744.11
MSD (x100-x96)	15793.48	16699.82	11185.94
MSD (x100-x94)	16295.20	16752.21	10642.99

Note: In each panel, x100, for instance, indicates parameter estimates of x(*a*, *b*, or *c*) when the number of observations is 100.

instability in Table 3 is not due to just several extreme values. Rather the instability is pervasive and substantial from one replication to another. Both mean absolute deviations and mean squared deviations are also very large indeed.

In the next series of Monte Carlo experiments, the parameter stability of irreducible cointegration is to be established. To this end, the following model is used for the DGP.

$$\begin{aligned} y_{1t} &= \beta_1 z_{1t}+\beta_2 z_{2t}+\varepsilon_{1t}, \\ y_{2t} &= z_{1t}+\varepsilon_{2t}, \\ y_{3t} &= z_{2t}+\varepsilon_{3t}, \end{aligned} \tag{11}$$

where each of ε_t's is iid series. The three series of ε_t's, with the standard deviations of 5.0, are separately generated to make them mutually uncorrelated. Two additional iid normal deviates with the standard deviations of 10.0, ε_{4t} and ε_{5t}, are generated with which

TABLE 5

STABLE COINTEGRATION RELATIONS IN IRREDUCIBLE COINTEGRATIONS

True Model: $y_{1t}=1.5z_{1t}+0.8z_{2t}+\varepsilon_{1t}$; $y_{2t}=z_{1t}+\varepsilon_{2t}$; and $y_{3t}=z_{2t}+\varepsilon_{3t}$.
Estimation: $y_{1t}-a-by_{2t}-c\ y_{3t}=\mathrm{I}(0)$. Expected values are $a=0.0$, $b=1.5$, and $c=0.8$.

Panel *A*: MLE results

	(1)	(2)	(3)	(4)
NOB	100	95	90	85
a	-0.0530	-0.0547	-0.0417	-0.0406
	(2.518)	(2.600)	(2.710)	(2.861)
b	1.5005	1.5005	1.5006	1.5008
	(0.040)	(0.043)	(0.047)	(0.052)
c	0.8000	0.8000	0.7996	0.7995
	(0.040)	(0.043)	(0.047)	(0.052)

Panel *B*: Correlation coefficients

a(NOB)	a(100)	a(95)	a(90)	a(85)
a(100)	1.00000	0.96619	0.93037	0.87477
		[0.0001]	[0.0001]	[0.0001]
a(95)		1.00000	0.96444	0.90838
			[0.0001]	[0.0001]
a(90)			1.00000	0.94420
				[0.0001]
b(NOB)	b(100)	b(95)	b(90)	b(85)
b(100)	1.00000	0.93355	0.85919	0.78353
		[0.0001]	[0.0001]	[0.0001]
b(95)		1.00000	0.92551	0.84140
			[0.0001]	[0.0001]
b(90)			1.00000	0.90614
				[0.0001]
c(NOB)	c(100)	c(95)	c(90)	c(85)
c(100)	1.00000	0.93294	0.86429	0.78448
		[0.0001]	[0.0001]	[0.0001]
c(95)		1.00000	0.92718	0.84186
			[0.0001]	[0.0001]
c(90)			1.00000	0.91335
				[0.0001]

Note: Values in parentheses are standard deviations over 10,000 replications and those in brackets are p-values to test if correlation coefficients are zero.

two I(1) factors of z_{1t} and z_{2t} are obtained by

$$\begin{aligned} z_{1t} &= 1.0 + z_{1t-1} + \varepsilon_{4t}, \\ z_{2t} &= 1.0 + z_{2t-1} + \varepsilon_{5t}, \end{aligned} \tag{12}$$

where the initial values are, as before, set to zero. Once 100 observations of z_t's and ε_t's are thus generated, three y_t's of 100 observations are computed through (11). By construction, there isonly one cointegration relation. That is, those three y_t's constitute the irreducible cointegration in this experiment. As before, each of y_t's is checked to make sure to be indeed I(1). The cointegration relation is then estimated by the MLE by using the lag length of one. The experiments are repeated 10,000 times, and each replication starts with new sets of normal deviates. In this way, all 10,000 experiments are made to be independent of one another.

Since there is only one cointegration relation, it is uniquely determined. Since β_1 is set to 1.5 and β_2 is set to 0.8, we expect that y_{1t}=constant$+1.5y_{2t}+0.8y_{3t}$ is the true cointegration relation. In order to discuss the parameter estimate stability, four different numbers of observations are used. But in this case, the following four observations are used: 100, 95, 90, and 85. Therefore, the degree of elimination of the last few observations is much severer here than in the earlier experiment, in which the numbers of observations are 100, 98, 96, and 94. Regardless of the observation numbers, the three time series are cointegrated in each of all 10,000 replications. The results from the MLE are provided in Table 5.

Cointegration relations are well defined for the irreducible cointegration. The estimate of a is expected to be zero, that of b is expected to be 1.5, and that of c is expected to be 0.8, because those values are used in the DGP. The MLE estimates in Panel A show that those true values are estimated very precisely. Values in parentheses are standard deviations over the 10,000 replications. The standard deviations for b and c are small indicating that the true values of 1.5 and 0.8 are indeed precisely estimated. Perhaps more importantly, the estimates are not sensitive to the number of observations used in the investigation. In Panel B, simple Pearson correlation coefficients are computed over the same 10,000 replications, for instance, between b from 100 observations and b from

TABLE 6

DIFFERENCES IN THE PARAMETER ESTIMATES IN IRREDUCIBLE COINTEGRATIONS

True Model: $y_{1t}=1.5z_{1t}+0.8z_{2t}+\varepsilon_{1t}$; $y_{2t}=z_{1t}+\varepsilon_{2t}$; and $y_{3t}=z_{2t}+\varepsilon_{3t}$.
Estimation: $y_{1t}-a-by_{2t}-cy_{3t}=\mathrm{I}(0)$.

Panel *A*: Mean deviations			
	a	b	c
Mean deviations (x100-x95)	0.0017	0.000044	0.000044
Mean deviations (x100-x90)	-0.0113	-0.000019	0.000370
Mean deviations (x100-x85)	-0.0124	-0.000290	0.000500
Panel *B*: Mean absolute deviations (MAD)			
	a	b	c
MAD (x100-x95)	0.4170	0.0095	0.0094
MAD (x100-x90)	0.6229	0.0148	0.0145
MAD (x100-x85)	0.8183	0.0198	0.0195
Panel C: Mean squared deviations (MSD)			
	a	b	c
MSD (x100-x95)	0.4494	0.00024	0.00024
MSD (x100-x90)	0.9869	0.00058	0.00056
MSD (x100-x85)	1.9218	0.00100	0.00100

Note: In each panel, x100, for instance, indicates parameter estimates of x(*a*, *b*, or *c*) when the number of observations is 100.

90 observations, which is 0.93355. This is significantly different from zero with the p-value of 0.0001. As expected, the correlation coefficient is the smallest when it is computed between 100 observations and 85 observations. The smallest of all correlation coefficient is 0.78353 for *b* nevertheless. Those values in Table 5 should be compared to the entries in Table 3 in order to contrast the stability in Table 5 and the instability in Table 3.

As before, mean deviations, mean absolute deviations, and mean squared deviations are reported in Table 6. All these measures indicate that the cointegration relations in this experiment are very stable. Especially for *b* and *c*, coefficients for I(1) time series, the measures in Table 6 show that cointegration relations are very stable when five, ten, or even 15 observations are removed from the 100 observations. The cointegration relations are obtained to be close to the true values with little variations. More importantly, they are not very sensitive to the changes in the number of observations.

VI. Conclusions

When a system of I(1) time series is cointegrated, there is a possibility that there are more than one cointegration relation. Suppose there are r cointegration relations among k I(1) time series. Then the number of common I(1) factor must be k-r. If a simultaneous equations model is fully provided, then the fact that there are k-r common factors and there are r cointegration relations is not very informative, because the given model specifies how each of I(1) endogenous variables is exactly determined by I(1) and I(0) exogenous variables. The inquiry into the cointegration becomes important when such a simultaneous equations model is not fully specified, and those common I(1) factors therefore are necessarily assumed to be latent.

When there is a multiple cointegration, however, the cointegration relations are not well defined, because there is an infinite number of cointegration relations possible. In this paper, we show that the presence of multiple cointegrations should lead to a construction of irreducible cointegration, in which the minimum set of I(1) series is searched to be cointegrated. A removal of any I(1) series from an irreducible cointegration will make the system no longer cointegrated. Moreover, once a set of I(1) series is cointegrated, an addition of any, whether related or unrelated, I(1) series will make the extended set also cointegrated, because the weight to the additional I(1) series can take a value of zero to maintain the set cointegrated.

The use the long-maturity interest rates in the United States shows the importance of estimating the irreducible cointegration relations. The four long-maturity rates of five-year, seven-year, ten-year, and 30-year, there is a multiple cointegration such that any three out of the four interest rates are also cointegrated. More importantly, the multiple cointegration relations are unreliable and unstable. The relations obtained from the MLE are very sensitive to a small change in the number of observations. When a few observations are removed from 216 observations, the cointegration relations substantially change. Although the statistical discrepancies are not tested for their significance, parameter estimates change substantially. On the other hand, the cointegration relations in the irreducible cointegration are very stable. In the case of the long-term interest rates, cointegration should be discussed in terms of three out of four rates. The instability of cointegration relations in multiple cointegrations is similar to that of regression coefficients in multiple regression equations. One standard way to resolve the difficulty of multicollinearity is to drop one or more variables from regressors (Greene 2000, p. 258). By dropping one or more I(1) series from multiple cointegrations to achieve irreducible cointegrations is also very similar to the solution for the multicollinearity.

In order to show the nature and the extent of the instability of multiple cointegrations, two sets of Monte Carlo experiments are conducted. As expected, parameter estimates are very unstable and unreliable from multiple cointegrations, whereas they are stable from irreducible cointegrations. When a given set of I(1) time series are cointegrated, analysts should identify and estimate the irreducible cointegration relations, because multiple cointegration relations are not uniquely defined and therefore they are unstable.

(*Received 28 November 2002; Revised 18 December 2002*)

References

Ahn, S. K., and Reinsel, G. C. "Estimation for Partially Nonstationary Multivariate Autoregressive Models." *Journal of American Statistical Association* 85 (No. 411 1990): 813-23.

Arshanapalli, Bala, and Doukas, John. "Common Stochastic Trends in a System of Eurocurrency Rates." *Journal of Banking and Finance* 18 (No. 6 1994): 1047-61.

Bowe, Michael, and Mylonidis, Nikolaos. "Is the European Capital Market Ready for the Single Currency?" *Journal of Business, Finance and Accounting* 26 (Nos. 1/2 1999): 1-32.

Bradley, Michael G., and Lumpkin, Stephen A. "The Treasury Yield Curve as a Cointegrated System." *Journal of Financial and Quantitative Analysis* 27 (No. 3 1992): 449-63.

Choi, Seungmook, and Wohar, Mark E. "New Evidence Concerning the Expectations Theory for the Short End of the Maturity Spectrum." *The Journal of Financial Research* 14 (No. 1 1991): 83-92.

_________. "The Expectations Theory of Interest Rates: Cointegration and Factor Decomposition." *International Journal of Forecasting* 11 (No. 2 1995): 253-62.

Cuthbertson, Keith, Hayes, Simon, and Nitzsche, Dirk. "Interest Rates in Germany and the UK: Cointegration and Error Correction Models." *The Manchester School* 66 (No. 1 1998): 27-43.

Engle, R. F., and Granger, C. W. J. "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55 (No. 2 1987): 251-76.

Engsted, Tom, "The Term Structure of Interest Rates in Denmark 1982-9: Testing the Rational Expectations/Constant Liquidity Premium-Theory." *Bulletin of Economic Research* 45 (No. 1 1993): 19-37.

Engsted, Tom, and Tanggaard, Carsten. "A Cointegration Analysis of Danish Zero-Coupon Bond Yields." *Applied Financial Economics* 4 (No. 4 1994): 265-78.

Granger, C. W. J. "Developments in the Study of Cointegrated Variables." *Oxford Bulletin of Economics and Statistics* 48 (No. 3 1986): 213-28.

_________. "Some Recent Generalizations of Cointegration and the Analysis of Long-Run Relationships." In R. F. Engle, and C. W. J. Granger (Eds.), *Long-Run Economic Relationships: Readings in Cointegration*. New York: Oxford University Press, 1991.

Granger, C. W. J., and Lee, Tae-Hwy. "Multicointegration." In Thomas B. Fomby, and George F. Jr. Rhodes (Eds.), *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots* (*Vol. 8*). Greenwich, Connecticut: JAI Press, pp. 71-84, 1990.

Granger, C. W. J., and Newbold, P. "Spurious Regression in Econometrics." *Journal of Econometrics* 2 (No. 2 1974): 111-20.

Greene, William H. *Econometric Analysis.* 4th Edition, Upper Saddle River, New Jersey: Prentice Hall, 2000.

Guest, Ross, and McLean, Alan. "New Evidence on the Expectations Theory of the Term Structure of Australian Commonwealth Government Treasury Yields." *Applied Financial Economics* 8 (No. 1 1998): 81-7.

Hamilton, James D. *Time Series Analysis*, Princeton, New Jersey: Princeton University Press, 1994.

Hiraki, Takato, Shiraishi, Noriyoshi, and Takezawa, Nobuya. "Cointegration, Common Factors, and the Term Structure of Yen Offshore Interest Rates." *The Journal of Fixed Income* 6 (No. 4 1996): 69-75.

Johansen, S. "Statistical Analysis of Cointegrating Vectors." *Journal of Economic Dynamics and Control* 12 (Nos. 2/3 1988): 231-54.

__________. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59 (No. 6 1991): 1551-80.

Johansen, S., and Juselius, K. "Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money." *Oxford Bulletin of Economics and Statistics* 52 (No. 2 1990): 169-210.

Luukkonen, Ritva, Ripatti, Antti, and Saikkonen, Pentti. "Testing for a Valid Normalization of Cointegration Vectors in Vector Autoregressive Processes." *Journal of Business and Economic Statistics*, 17 (No. 2 1999): 195-204.

Mandeno, Robert J., and Giles, David E. A. "The Expectations Theory of the Term Structure: A Cointegration/Causality Analysis of US Interest Rates." *Applied Financial Economics* 5 (No. 5 1995): 273-83.

McDermott, C. John. "Testing the Expectations Model of the Term Structure in Times of Financial Transition." *Applied Financial Economics* 8 (No. 6 1998): 663-9.

McFadyen, James, Pickerill, Karen, and Devany, Mike. "The Expectations Hypothesis of the Term Structure: More Evidence." *Journal of Economics and Business* 43 (No. 1 1991): 79-85.

Rossana, Robert J. "On the Adjustment Matrix in Error Correction Models." *Journal of Monetary Economics* 42 (No. 2 1998): 427-44.

Shea, Gary S. "Benchmarking the Expectations Hypothesis of the Interest-Rate Term Structure: An Analysis of Cointegration Vectors." *Journal of Business and Economic Statistics* 10 (No. 3 1992): 347-66.

Stock, James H., and Watson, Mark W. "Testing for Common Trends." *Journal of the American Statistical Association* 83 (No. 404 1988): 1097-107.

Wallace, Myles S., and Warner, John T. "The Fisher Effect and the

Term Structure of Interest Rates: Tests of Cointegration." *Review of Economics and Statistics* 75 (No. 2 1993): 320-4.

Warne, Anders. "Inference in Cointegrated VAR Systems." *Review of Economics and Statistics* 79 (No. 3 1997): 508-11.

Wickens, Michael R. "Interpreting Cointegrating Vectors and Common Stochastic Trends." *Journal of Econometics* 74 (No. 2 1996): 255-71.