Fractionality and co-fractionality between Government Bond yields: Implications for the yield curve

Håvard Hungnes*
Research Department
Statistics Norway
P.O.B. 8131 Dep
N-0033 Oslo
Norway

February 16, 2016

Abstract

In a co-fractional vector autoregressive (VAR) model two more parameters are estimated, compared to the traditional cointegrated VAR model. The increased number of parameters that needs to be estimated leads to identification problems; there is no unique formulation of a co-fractional system, though usually one formulation is preferred. This paper has the following contributions: (i) it discusses different kinds of identification problems in co-fractional VAR models; (ii) it proposes a specification test for higher order fractional processes; (iii) it presents an Ox program that can be used for estimating and testing co-fractional systems; and (iv) it uses the above mentioned contributions to analyse a system of Government Bonds in the US and Norway where the results indicates that the level and trend in the yield curve have a longer memory than the curvature (i.e., a linear combination of the yields of the Government Bonds that corresponds to representing the curvature of the yield curve is a co-fractional relationship).

Keywords: Fractional cointegration

Acknowledgments: Thanks to Bruce Hansen and Anders Rygh Swensen for valuable comments on an earlier version of this paper.

*E-mail: hhu@ssb.no. Homepages: http://people.ssb.no/hhu and http://www.hungnes.net

1 Introduction

In a co-fractional vector autoregressive (VAR) model two more parameters are estimated, compared to the traditional cointegrated VAR model (Johansen , 1988, 1996). These two additional parameters, usually denoted d and b, describes fractionally integrated orders in the system. The parameter d denotes the overall fractional integrated order of the system and b denotes the reduced fractional order of integration in the co-fractional relationships. The latter parameter also express the lag length used in the formulation, see Johansen (2008).

One interesting type of identification problem arises when there are no co-fractional relationships among the variables. This system can be formulated in two different ways in a co-fractional system. First, it can be formulated with a full co-fractional rank where the system as a whole is fractionally integrated of order $d + b$ but where the full rank implies that all the variables in the system is fractionally integrated of order d. Second, it can be formulated with a co-fractional rank of zero where the whole system is fractionally integrated of order d. The two formulations are just reparameterizations of each other, and therefore equivalent formulations, where the latter formulation needs one extra lag. This is also shown in Carlini and Santucci de Magistris (2013).

This equivalence result has an important implication for co-fractional rank testing. For the test of the null hypothesis of a rank equal to zero with a full rank as the alternative (i.e., a trace test for co-fractional rank), the equivalence result implies that this is closely related to a lag-reduction test. With one formulation of the parameter space, the co-fractional rank test of zero rank is always identical to a lag-reduction test. This identification problem is shown as a special case of a more general identification problem. In this paper I propose a slightly different parameter space for the fractional parameters d and b than in Johansen (2008).

The more general type of identification problems arise when too many lags are included in the specification. Two such identification problems are presented here; one which is also presented in Carlini and Santucci de Magistris (2013) and another that is — to my knowledge — not shown before. This is the first novel contribution of the paper and shown in Section 2.

Carlini and Santucci de Magistris (2013) suggest applying a so-called $F(d)$ condition for testing for such identification problems, where they suggest to simulate the distribution of their test statistic for each time it is applied. Here I suggest to reformulate the system in as in Johansen (1996) for $I(2)$ systems and test for fractionality at a higher order of integration. The test suggested here builds on the same idea as

Carlini and Santucci de Magistris (2013), but the main advantage is that the critical values of the test are available. Hence, additional simulation for critical values are not necessary. This is the second contribution of this paper, and is shown in Section 3.

The paper accommodate an Ox program for estimating co-fractional systems. This program can be used interactively and makes estimation of co-fractional systems easy. The Ox program is the third contribution of this paper and is presented in Section 4.

The paper also shows how to apply co-fractional systems to test for long memory in Government bond yields. The estimation of these systems also shows the identification problems in practice; both with respect to how this can lead to erroneous estimation of the co-fractional parameters and how to deal with these identification problems in order to end up with a formulation of the system that is interpretable. The analysis of the yield curves for the US and Norway is the forth contribution of the paper and is presented in Section 5.

2 Formulation of co-fractional systems

In order to understand the concept of fractional and co-fractional processes means, consider the univariate process $\Delta^d y_t = e_t$ where e_t is i.i.d. and Δ^d is the fractional difference operator. The fractional difference operator is defined by the binomial expansion $\Delta^d y_t = \sum_{n=0}^{\infty}(-1)^n\binom{d}{n}y_{t-n} = y_t + \sum_{i=1}^{\infty}\frac{\prod_{i=0}^{n-1}(-d+i)}{n!}y_{t-n}$. This variable is integrated of order d, also denoted $y_t \sim I(d)$, where $d > 0$. If $d = 1$ the fractional difference operator simplifies to the standard difference operator, Δ. If two variables are fractionally integrated of order d, they can be co-fractional integrated of order $d - b$ (where $d - b \geq 0$), i.e., $(y_{1t} - \zeta y_{2t}) \sim I(d - b)$ or $\Delta^{d-b}(y_{1t} - \zeta y_{2t}) \sim I(0)$. Standard cointegration is a special case of fractional co-fractional cointegration where $d = b = 1$.

Johansen (2008) suggest the following formulation of the co-fractional vector autoregressive (VAR) model for the n variables in Y_t as

$$\Delta^d Y_t = \Pi L_b \Delta^{d-b} Y_t + \sum_{i=1}^{\ell}\Gamma_i L_b^i \Delta^d Y_t + \varepsilon_t, \tag{1}$$

where Δ^d is the fractional difference operator and $L_b = 1 - \Delta^b$ is the fractional lag operator. The coefficient matrices Γ_i $(i = 1, \ldots, \ell)$ are of dimension $n \times n$, and the $n \times n$ coefficient matrix Π can have reduced rank and, hence, we define $\Pi = \alpha\beta'$ where both α and β are of dimension $n \times r$ and have full column rank such that $\text{rank}(\Pi) = r$. Finally, ε_t is i.i.d. $(0, \Omega)$ in n dimensions.

Let $\alpha_\perp$ and $\beta_\perp$ be the orthogonal complement of α and β, respectively, both with

dimension $n \times (n-r)$.[1] Furthermore, let $\Gamma = \sum_{i=1}^{\ell} \Gamma_i - I_n$. For the system in (1) not to be integrated of an higher order than d the following assumptions must hold when $r < n$:

Assumption 2.1 *The $(n-r) \times (n-r)$ matrix $\alpha_\perp' \Gamma \beta_\perp$ has full rank.*

This is an assumption known from both standard cointegrated VAR models (see, e.g., Johansen , 1996) and co-fractional VAR models (Johansen and Nielsen , 2012). Carlini and Santucci de Magistris (2013) show that this is a sufficient condition for identification when $n > r > 0$. Furthermore, they suggest how one can apply this criteria in practise. In Section 3 I suggest an alternative approach for testing this assumption that does not involve additional simulations.

Furthermore, I will assume the following assumption is satisfied:

Assumption 2.2 *The coefficient matrix $\Gamma_\ell \neq 0$.*

This assumption is included to avoid the type of equivalent formulation for any co-fractional order, as one can always add one extra lag in the system where the corresponding coefficient matrix consists of zeros only.

In Section 2.1 I will consider two types of identification problems. In addition, I describe the general consequence of each of the two types of identification problems. In Section 2.2 I discuss the parameter spaces for the two fractional parameter and the implications of these parameter spaces on the identification problems. In Section 2.3 I show how one of the identification problems always implies that a test of a co-fractional rank is equal to zero versus the alternative of a full co-fractional rank is identical to a lag reduction test for a particular choice of parameter set.

2.1 Two identification problems - illustrations

To illustrate the identification problems it will be convenient to rewrite the system as

$$\Delta^{d+b} Y_t = \Pi L_b \Delta^{d-b} Y_t + \Gamma L_b \Delta^d Y_t + \sum_{i=1}^{\ell-1} \Gamma_i^* L_b^i \Delta^{d+b} Y_t + \varepsilon_t, \tag{2}$$

where $\Gamma_i^* = -\sum_{j=i+1}^{\ell} \Gamma_j$ $(i = 1, \ldots, \ell - 1)$.

Now consider a simple version of (1) with no lags in differences (i.e., $\ell_1 = 0$);

$$\Delta d^{(1)} Y_t = \Pi^{(1)} L_{b^{(1)}} \Delta^{d^{(1)} - b^{(1)}} Y_t + \varepsilon_t, \tag{3}$$

[1] The orthogonal complement of the full column rank matrix A is denoted $A_\perp$ and is defined such that $A_\perp' A = 0$ and $(A, A_\perp)$ has full rank. (The orthogonal complement of a nonsingular matrix is 0, and the orthogonal complement of a zero matrix is an identity matrix of suitable dimension.)

where the super script (1) is used to denote one particular parameter set; here parameter set 1.

2.1.1 Identification problem 1 - illustration

One reformulation of his system (1) or (2) where $\ell^{(2)} = 1$;

$$\begin{aligned} \Delta^{d^{(2)}} Y_t &= \Pi^{(2)} L_{b^{(2)}} \Delta^{d^{(2)}-b^{(2)}} Y_t + \Gamma_1^{(2)} L_{b^{(2)}} \Delta^{d^{(2)}} Y_t + \varepsilon_t \\ &\Updownarrow \\ \Delta^{d^{(2)}+b^{(2)}} Y_t &= \Pi^{(2)} L_{b^{(2)}} \Delta^{d^{(2)}-b^{(2)}} Y_t + \Gamma^{(2)} L_{b^{(2)}} \Delta^{d^{(2)}} Y_t + \varepsilon_t \end{aligned} \tag{4}$$

This system in (4) is equivalent with (3) when $b^{(2)} = b^{(1)}$, $d^{(2)} = d^{(1)} - b^{(1)}$, $\Pi^{(2)} = 0$, and $\Gamma^{(2)} = \Pi^{(1)}$ (where the latter implies $\Gamma_1^{(2)} = I_n + \Pi^{(1)}$).

First, note that if $\Pi^{(1)}$ has reduced rank, then Assumption 2.1 does not hold for parameter set 2. In this case we have $\alpha_\perp^{(2)} = \beta_\perp^{(2)} = I_n$ with parameter set 2 (since $\alpha^{(2)} = \beta^{(2)} = 0$ follows from $\Pi^{(2)} = 0$). Therefore; the criteria in the assumption becomes $(\alpha_\perp^{(2)})'\Gamma^{(2)}\beta_\perp^{(2)} = \Gamma^{(2)} = \Pi^{(1)}$, which has reduced rank. Hence, the criteria in Assumption 2.1 is not fulfilled for parameter set 2 when Π has reduced rank.

Second, note that if $\Pi^{(1)}$ has full rank, Assumption 2.1 holds for both parameter set 1 and parameter set 2. With parameter set 1 the system is formulated as a system with full co-fractional rank, whereas for parameter set 2 the system is formulated with a co-fractional rank of zero. This will be shown below (in Corollary 2.1) to be a general result: a system with full co-fractional rank with any number of lags can also be formulated as a system with a co-fractional rank of zero, where the latter system has more lags than the former. Note also that — given the co-fractional rank and the lag length — the number of free parameters are the same in the two formulations. Therefore, the likelihood value will also be identical for the two formulations.

Now I will show this result more generally. Let $\lambda^{(1)} = \lambda^{(1)}(\ell^{(1)}, r^{(1)}) = (d^{(1)}, b^{(1)}\Pi^{(1)}, \Gamma_1^{(1)}, \ldots, \Gamma_{\ell^{(1)}}^{(1)}, \Omega^{(1)})$ be parameter set 1 with the reformulation of the co-fractional system in (1) for any lag length $\ell \geq 0$. The parameter set 2 is defined similarly.

Proposition 2.1 (The non-uniqueness involving higher order integration I) *Consider the system in* (1) *with parameter set 1, i.e.,* $\lambda^{(1)} = \lambda^{(1)}(\ell^{(1)}, r^{(1)}) = (d^{(1)}, b^{(1)}\Pi^{(1)}, \Gamma_1^{(1)}, \ldots, \Gamma_{\ell^{(1)}}^{(1)}, \Omega^{(1)})$. *Furthermore, let k be any positive integer. An equivalent formulation of this system is* (1) *with the following parameter set;* $\lambda^{(2)} = \lambda^{(2)}(\ell^{(2)}, r^{(2)}) = (d^{(2)}, b^{(2)}, \Pi^{(2)}, \Gamma_1^{(2)}, \ldots, \Gamma_{\ell^{(2)}}^{(2)}, \Omega^{(2)})$ *with* $d^{(2)} = d^{(1)} - kb^{(1)}$, $b^{(2)} = b^{(1)}$, $\ell^{(2)} = \ell^{(1)} + k$,

and $\Pi^{(2)} = 0$ *which implies* $r^{(2)} = 0$*. If* $k = 1$*:* $\Gamma^{(2)} = \Pi^{(1)}$*, which implies that* $\alpha'_\perp \Gamma \beta_\perp$ *has reduced rank when* $r < n$*. If* $k > 1$*:* $\Gamma^{(2)} = 0$*, i.e., the criteria in Assumption 2.1 has reduced rank for all* r*. Finally;* $\Omega^{(2)} = \Omega^{(1)}$*, hence, the two formulations are just reparametrizations of each other.*

The proof is given in the Appendix. A consequence of this proposition is the following result:

Corollary 2.1 *If* $\Pi^{(1)}$ *has full rank and Assumption 2.1 holds for parameter set 1, then — for* $k = 1$ *— we have* $\Pi^{(2)} = 0$ *and* $\ell^{(2)} = \ell^{(1)} + 1$ *under parameter set 2 and for this particular k Assumption 2.1 holds also for parameter set 2.*

2.1.2 Identification problem 2 - illustration

The second type of identification problem follows Carlini and Santucci de Magistris (2013). Now consider (1) or (2) with $\ell_3 = 1$;

$$\begin{aligned} \Delta^{d^{(3)}} Y_t &= \Pi^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t + \Gamma_1^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}} Y_t + \varepsilon_t \\ &\Updownarrow \\ \Delta^{d^{(3)}+b^{(3)}} Y_t &= \Pi^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t + \Gamma^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}} Y_t + \varepsilon_t \end{aligned} \tag{5}$$

Furthermore, to show the identification problem, use

$$\begin{aligned} & \Pi^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t \\ = \; & \Pi^{(3)} L_{2b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t + \Pi^{(3)} L_{b^{(3)}} \left(1 - L_{b^{(3)}}\right) \Delta^{d^{(3)}-b^{(3)}} Y_t \\ = \; & \Pi^{(3)} L_{2b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t + \Pi^{(3)} L_{b^{(3)}} \Delta^{b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t \\ = \; & \Pi^{(3)} L_{2b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t + \Pi^{(3)} L_{b^{(3)}} \Delta^{d^{(3)}} Y_t. \end{aligned}$$

Inserting this into (5) yields:

$$\begin{aligned} \Delta^{d^{(3)}+b^{(3)}} Y_t &= \Pi^{(3)} L_{2b^{(3)}} \Delta^{d^{(3)}-b^{(3)}} Y_t \\ &\quad + \left(\Gamma^{(3)} + \Pi^{(3)}\right) L_{b^{(3)}} \Delta^{d(3)} Y_t + \varepsilon_t \end{aligned} \tag{6}$$

Now we can see that the formulation in (3) is equivalent with (5) with the following parameter restrictions: $b^{(3)} = \frac{1}{2} b^{(1)}$, $d^{(3)} = d^{(1)} - \frac{1}{2} b^{(1)}$, $\Pi^{(3)} = \Pi^{(1)}$, and $\Gamma^{(3)} + \Pi^{(3)} = 0$ (where the latter implies $\Gamma_1^{(3)} = I_n - \Pi^{(1)}$).

Note that if $\Pi^{(1)}$ has reduced rank, then Assumption 2.1 does not hold for parameter set 3. With parameter set 3 and if $0 < r < n$ we have $\alpha_\perp^{(3)} = \alpha_\perp^{(1)}$, $\beta_\perp^{(3)} = \beta_\perp^{(1)}$

and $\Gamma^{(3)} = -\Pi^{(1)} = -\alpha^{(1)}\beta^{(1)\prime}$.[2] Applying this in the criteria in Assumption 2.1 yields $\alpha_\perp^{(3)\prime}\Gamma^{(3)}\beta_\perp^{(3)} = \alpha_\perp^{(1)\prime}\alpha^{(1)}\beta^{(1)\prime}\beta_\perp^{(1)} = 0$, which has a rank equal to zero, and — hence — a reduced rank. If $r = 0$, the criteria in Assumption 2.1 is the rank of $\Gamma^{(3)}$, which clearly has a rank equal to zero since $\Gamma^{(3)} = \Pi^{(0)} = 0$.

If $\Pi^{(1)}$ has full rank, the criteria in Assumption 2.1 is not defined and can not be used to choose between the two formulations.

A general formulation of this identification problem is given in the following proposition:

Proposition 2.2 (The non-uniqueness involving higher order integration II) *Consider the system in* (1) *with transformed parameter set 1, i.e.,* $\lambda^{(1)} = \lambda^{(1)}(\ell^{(1)}, r^{(1)}) = (d^{(1)}, b^{(1)}\Pi^{(1)}, \Gamma_1^{(1)}, \ldots, \Gamma_{\ell^{(1)}}^{(1)}, \Omega^{(1)})$. *Furthermore, let k be any positive integer exceeding one. An equivalent formulation of this system is* (1) *with the following transformed parameter set;* $\lambda^{(3)} = \lambda^{(3)}(\ell^{(3)}, r^{(3)}) = (d^{(3)}, b^{(3)}, \Pi^{(3)}, \Gamma_1^{(3)}, \ldots, \Gamma_{\ell^{(3)}}^{(3)}, \Omega^{(3)})$ *with* $d^{(3)} = d^{(1)} - \frac{k-1}{k}b^{(1)}$, $b^{(1)} = b^{(1)}/k$, $\ell^{(3)} = (k+1)\ell^{(1)} - 1$, $\Pi^{(3)} = \Pi^{(1)}$ *(so* $r^{(3)} = r^{(1)}$*). Furthermore,* $\Gamma^{(3)} = -\Pi^{(1)}$ *for any k, so the criteria in Assumption 2.1 has reduced rank for any feasible k as long as* $r < n$. *Finally;* $\Omega^{(3)} = \Omega^{(1)}$, *hence, the two formulations are just reparametrizations of each other.*

The proof is given in Carlini and Santucci de Magistris (2013) and repeated in the Appendix.

2.2 Parameter space and true parameters

The difference operator Δ^d is well-defined for non-negative values of d. For $d = 0$ we have $\Delta^0 Y_t = Y_t$. Hence, the left hand side of (1) is well-defined for $d \geq 0$. Similarly, the co-fractional term is well defined for $d - b \geq 0$. The lag-operator $L_b = 1 - \Delta^b$ is not defined for $b < 0$. Furthermore, $b = 0$ implies $L_0 = 0$ which is equivalent with no lags and, hence, that the parameter coefficients Γ_i $(i =, \ldots, \ell)$ are not identifiable. Therefore, b is restricted to be strictly positive. Johansen and Nielsen (2012) considers the following parameter space:

Parameter space 1 *The parameter space for d and b is given by* $d \geq b > 0$ *if* $r > 0$ *or* $\ell > 0$. *For* $\ell = r = 0$, *b is not a part of the system, and the parameter space for d is given by* $d \geq 0$.

[2] Actually, only the space spanned by α and $\alpha_\perp$ are identified and whereas the parameter matrices α and $\alpha_\perp$ are not. Therefore, to be precise we have $sp(\alpha_\perp^{(3)}) = sp(\alpha_\perp^{(1)})$ and $sp(\beta_\perp^{(3)}) = sp(\beta_\perp^{(1)})$ where $sp(\alpha)$ denotes space spanned by α. However, since only the space spanned by these matrices are identifiable, we can always choose a normalisation such that $\alpha_\perp^{(3)} = \alpha_\perp^{(1)}$ and $\beta_\perp^{(3)} = \beta_\perp^{(1)}$.

However, for co-fractional rank equal to zero, the term involving the difference operator Δ^{d-b} is not a part of the system. Hence, the parameter space can be extended for $r = 0$:

Parameter space 2 *The parameter space for d and b is given by (i) $d \geq b > 0$ if $r \geq 1$ and (ii) $d > 0$ and $b \geq 0$ for $r = 0$ and $\ell \geq 1$. For $\ell = r = 0$, b is not identifiable, and the parameter space for d is given by $d \geq 0$.*

The parameter space in Parameter space 1 implies that Corollary 2.1 does not hold for all parameters. If $2b_2 > d_2 > b_2$ under full rank, the corresponding parameters in the zero-rank formulation will imply $d_1 = d_2 - b_2 < b_2 = b_1 \Longrightarrow d_1 < b_1$, which is not feasible given the parameter space in Parameter space 1.

With the parameter space in Parameter space 2, however, Corollary 2.1 will be valid for all feasible parameters. For example, consider the parameters $d_2 \geq b_2 > 0$ under full rank, which corresponds to $d_1 = d_2 - b_1 \geq 0$ and $b_2 = b_1 > 0$ with zero-rank.

The equivalence results imply that it is not straightforward to define 'true parameters'. In particulare, the 'true value' of d differs between the full-rank formulation and the zero-rank formulation.

2.3 The co-fractional rank test as a lag-reduction test

For a given lag length, ℓ, a given rank, r, and given values of the fractional parameters d and b, the system in (1) can be estimated as standard cointegrated VAR models, as described in Johansen (1996) or Johansen and Nielsen (2012). Let the corresponding profile likelihood function where $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_\ell, \Omega)$ have been concentrated out to be denoted $L(d, b, \ell, r)$. Furthermore, let $L\left(\hat{\lambda}\left(\ell, r; d \geq d^{min}\right)\right) = \max_{d \geq d^{min}, b > 0} L(d, b, \ell, r)$, be the likelihood value when estimating for a lag length ℓ and a rank r where the parameter space for d is stated explicit. The parameter d^{min} take the value b for a co-fractional rank greater than one, and takes the value b or 0 for a co-fractional rank of zero, depending on the chosen parameter space (i.e., if Parameter space 1 or Parameter space 2 is chosen). Corollary 2.1 can now in terms of likelihood value be formulated as

$$L\left(\hat{\lambda}\left(\ell, 0; d \geq 0\right)\right) = L\left(\hat{\lambda}\left(\ell - 1, n; d \geq b\right)\right). \tag{7}$$

The likelihood ratio test for testing the null hypothesis of a co-fractional rank r against a full rank n is given by

$$LR(n - r) = 2 \log\left[L\left(\hat{\lambda}\left(\ell, n; d \geq b\right)\right) / L\left(\hat{\lambda}\left(\ell, r; d \geq d^{min}\right)\right)\right]. \tag{8}$$

First, consider this test with the parameter space given by Parameter space 2. Then $d^{min} = b$ if $r \geq 1$ and $d^{min} = 0$ if $r = 0$, where the interest here is on the case with $r = 0$. Applying (7) yields

$$LR(n) = 2\log\left[\frac{L\left(\hat{\lambda}\left(\ell, n; d \geq b\right)\right)}{L\left(\hat{\lambda}\left(\ell - 1, n; d \geq b\right)\right)}\right]. \tag{9}$$

This expression is identical to a lag-reduction test. Hence, when allowing the parameter set to be given by Parameter space 2, the co-fractional rank test of a rank equal to zero is identical to a lag-reduction test as long as $\ell > 0$.

Second, consider the co-fractional rank test with the parameter space given by Parameter space 1, which is the parameter space considered by Johansen and Nielsen (2012). Since this is a test with a smaller parameter space for the rank equal to zero, the test will not have the same distribution as above if the parameter restriction for d is binding.

3 Testing for higher order fractional integration

The formulation in (2) can be used to distinguish between two cases: (i) normal co-fractionality, and (ii) polynomial co-fractionality. Normal co-fractionality implies that the variables in Y are integrated of order d but there possibly exists one or more linear relationships between these variables that are integrated of order $d - b$. A simple type of polynomial co-fractionality implies here that Y is integrated of order $d + b$, and there are linear combinations of the variables in Y and $\Delta^b Y$ that are integrated of order $d - b$. In this case both Π and $\alpha'_\perp \Gamma \beta_\perp$ has reduced rank in (2).

First I will derive the necessary and sufficient restriction for the system to not involve polynomial co-fractionality. Here, I follow the exposition in Johansen (1996) for $I(2)$-systems but with the necessary adjustments to allow for co-fractionality. Assume that (α, β, r, d, b) are known (or estimated). Then also $\alpha_\perp$ and $\beta_\perp$ are known. The question is now if this implies normal co-fractionality, e.g., that Y is integrated of order d and $\beta' Y$ is integrated of order $d - b$. To derive the condition for normal co-fractionality, I pre-multiply (2) with $\alpha'_\perp$ and get

$$\alpha'_\perp \Delta^{d+b} Y_t = \alpha'_\perp \Gamma L_b \Delta^d Y_t + \sum_{i=1}^{\ell-1} \alpha'_\perp \Gamma_i^* L_b^i \Delta^{d+b} Y_t + \alpha'_\perp \varepsilon_t. \tag{10}$$

Furthermore, applying $I_n = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' + \beta(\beta'\beta)^{-1}\beta'$, we get

$$\begin{aligned}\left[\alpha_\perp'\Delta^{d+b}Y_t\right] &= \alpha_\perp'\Gamma\beta_\perp\left[(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'L_b\Delta^d Y_t\right] + \alpha_\perp'\Gamma\beta(\beta'\beta)^{-1}\left[\beta'L_b\Delta^d Y_t\right] \\ &\quad + \sum_{i=1}^{\ell-1}\alpha_\perp'\Gamma_i^*\left[L_b^i\Delta^{d+b}Y_t\right] + \alpha_\perp'\varepsilon_t. \qquad (11)\end{aligned}$$

For Y to be fractionally integrated of order d then $\alpha_\perp'\Gamma\beta_\perp$ must have full rank. If $\alpha_\perp'\Gamma\beta_\perp$ do not have full rank Y is integrated of order $d+b$ or higher.

Carlini and Santucci de Magistris (2013) suggest testing for full rank by testing if $|\alpha_\perp'\Gamma\beta_\perp| = 0$. If the hypothesis $|\alpha_\perp'\Gamma\beta_\perp| = 0$ is rejected $\alpha_\perp'\Gamma\beta_\perp$ has full rank, and — hence — Y is integrated of order d. If the hypothesis is not rejected, Carlini and Santucci de Magistris (2013) indirectly assumes that it has a rank equal to zero. They then suggest to reduce the lag length ℓ and reestimate the system.

One disadvantage with the approach suggested by Carlini and Santucci de Magistris (2013) is that the distribution of the proposed test must be simulated for each data set. Another disadvantage is that the test does not distinguish between when $\alpha_\perp'\Gamma\beta_\perp$ has a rank of zero or a reduced rank. These situations have two different implications. A reduced rank can indicate that the identification problem in Proposition 2.1 applies. It can also indicate that there are some sort of polynomial co-fractionality among the series in Y. However, if the rank is equal to zero there still can be a situation with normal co-fractionality, though the difference in the order of integration between Y and the cofractioanal relatioship is a multiplicity of b, see Carlini and Santucci de Magistris (2013). I suggest estimating (11) with reduced rank regression, where (α, β, r, d, b) is assumed known (and set equal to the estimates when estimating (1)). This is similar to the misspecification test for the presence of $I(2)$ suggested by Johansen (1996). The approach implies regressing $\alpha_\perp'\Delta^{d+b}Y_t$ on $(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'L_b\Delta^d Y_t$ corrected for $\beta'L_b\Delta^d Y_t$ and $L_b\Delta^{d+b}Y_t, \ldots, L_b^{\ell-1}\Delta^{d+b}Y_t$. This is a normal fractional system, and when testing for reduced rank critical test values for such systems can be used. Hence, we can use this formulation to test for wether $\alpha_\perp'\Gamma\beta_\perp$ has full rank or not.

4 Estimating co-fractional systems with `Cofrac` for Ox

`Cofrac` is an Ox Program for estimating co-fractional systems.[3] It is run as an OxPack program in Ox Professional, which makes it an interactive program with a easy-to-use

[3]The program can be downloaded from http://www.hungnes.net/cofrac. In requires OX Professional, see Doornik (2007) and http://www.oxmetrics.com.

graphical user interface. It is therefore easy to use to estimate co-fractional systems.

`Cofrac` can be used to estimate co-fractional systems such as in (1). The special case where the coefficient restriction $b = d$ is imposed, see Johansen and Nielsen (2012), can also be estimated. In this case an intercept is included in the co-fractional relationship:

$$\Delta^d Y_t = \Pi L_d Y_t^* + \sum_{i=1}^{\ell} \Gamma_i L_d^i \Delta^d Y_t + \varepsilon_t,$$

where $Y_t^* = (Y_t', 1)'$. Note that with this restriction the co-fractional relationship is $I(0)$.

`Cofrac` can also be used to estimate where all the variables in Y_t are integrated of different orders. See e.g. Tschering et al. (2013) for an example of such formulation of a co-fractional system. This specification can be considered as a generalization of a standard co-fractional system where a standard co-fractional analysis is conducted on a system where the difference in the order of integration for all variables in Y are corrected for.

$$\begin{pmatrix} \Delta^{d_1} y_{1,t} \\ \vdots \\ \Delta^{d_n} y_{n,t} \end{pmatrix} = \Pi L_b \begin{pmatrix} \Delta^{d_1-b} y_{1,t} \\ \vdots \\ \Delta^{d_n-b} y_{n,t} \end{pmatrix} + \sum_{i=1}^{\ell} \Gamma_i L_b^i \begin{pmatrix} \Delta^{d_1} y_{1,t} \\ \\ \Delta^{d_n} y_{n,t} \end{pmatrix} + \varepsilon_t$$

$$\Updownarrow$$

$$\Delta^{d_1} \begin{pmatrix} y_{1,t} \\ \Delta^{d_2-d_1} y_{2,t} \\ \vdots \\ \Delta^{d_n-d_1} y_{n,t} \end{pmatrix} = \Pi L_b \Delta^{d_1-b} \begin{pmatrix} y_{1,t} \\ \Delta^{d_2-d_1} y_{2,t} \\ \vdots \\ \Delta^{d_n-d_1} y_{n,t} \end{pmatrix} + \sum_{i=1}^{\ell} \Gamma_i L_b^i \Delta^{d_1} \begin{pmatrix} y_{1,t} \\ \Delta^{d_2-d_1} y_{2,t} \\ \vdots \\ \Delta^{d_n-d_1} y_{n,t} \end{pmatrix} + \varepsilon_t$$

In `Cofrac` it is also possible to set restrictions for the parameter space for b and d. For example, the lower restriction for the fractional parameter b can be set to 0 or $\frac{1}{2}$; the upper restriction to $\frac{1}{2}$ or infinite. Hence, it is possible to force the co-fractional rank test to be χ^2-distributed by imposing that $b < \frac{1}{2}$; or have a non-standard distribution by imposing $b > \frac{1}{2}$. `Cofrac` also allows for restrictions on $d - b$.[4] For example, the restriction $d - b = 0$, which implies that the co-fractional relationship is integrated of

[4]For systems with co-fractional rank equal to zero (i.e. no co-fractional relationships), the user can choose to put restrictions on d rather than $d - b$, i.e., whether to use Parameter set 1 or Parameter set 2.

order 0. With this restriction, an intercept is automatically included in the co-fractional relationships, see Johansen and Nielsen (2012). Alternatively the restriction $d - b < \frac{1}{2}$ can be imposed. This implies that the co-fractional relationships are strictly stationary, i.e. have both a mean and a finite variance.

For given values of the fractional parameters d and b the following vector of variables can be constructed[5]

$$Z_{0,t} = \Delta^d Y_t, Z_{1,t} = L_b \Delta^{d-b} Y_t, \text{ and } Z_{2,t} = \left(L_b \Delta^d Y_t', \ldots, L_b^\ell \Delta^d Y_t' \right)',$$

and the co-fractional system can be formulated as

$$Z_{0t} = \alpha \beta' Z_{1t} + \Gamma Z_{2t} + \varepsilon_t,$$

where $\Gamma = (\Gamma_1, \ldots, \Gamma_\ell)$. This system can be estimated as standard cointegrated systems, see e.g. Johansen (1996). In order to find the full information maximum likelihood estimates of all the parameters, i.e., also d and b, a simulation procedure must be applied. Since also `Cofrac` allows for imposing restriction on the co-fractional parameters d and b, the MaxSQP algorithm in Ox is applied. According to the Ox documentation, see Doornik (2007), MaxSQP implements a sequential programming technique to maximize a non-linear function subject to non-linear constraints, similar to Algorithm 18.7 in Nocedal and Wright (1999).

If d is not an integer, $\Delta^d X_t$ is a function of infinite number of lags of X_t. To construct a feasable counterpart of $\Delta^d X_t$ `Cofrac` construct a "type II" (truncated) fractional process as suggested by Johansen and Nielsen (2010). Hence, `Cofrac` applies $Z_{0t} = \sum_{i=0}^{N} (-1)^n \binom{d}{i} X_{t-i}$, where N is the number of initialising observations. In `Cofrac` the user can set the value of N.

The constructed variables $Z_{0,t}$, $Z_{1,t}$ and $Z_{2,t}$ are function of d and b and must be updated throughout the search algorithm for obtaining the maximum likelihood estimates. The updating of these variables are conducted as suggested in Jensen and Nielsen (2014) to speed up the updating of the constructed variables.

The search algorithm for obtaining the estimates of the parameters need some starting values for the fractional parameters. The identification problems presented in Section 2 indicates that the likelihood function may not be well-behaved. Hence, good starting values can be essential to reach a global maximum likelihood, and not just a local maximum likelihood. To get good staring values the program starts out with a grid search over the parameters b and d. For both b and $d - b$ the grid involves 11

[5] If the restriction $d - b = 0$ is imposed, an intercept is included in Z_{1t}.

different values from 0 to 2 (where the lower and upper limit is changed if it is outside the allowed parameter space).[6]

In `Cofrac` one can also estimate systems where restrictions on the co-fractional relationships are imposed as $\beta = H'\phi$ where H is an $n \times s$ matrix of specified parameters where $n < s < 0$ such that ϕ is $s \times r$ of unknown parameters that are to be estimated.

`Cofrac` also test for higher order fractional integration as suggested in Section 3.

5 Government bonds and the yield curve

Factor models are popular for modelling the yield of bonds. According to Duffee (2011) and Diebold and Rudebusch (2013) only a few factors are sufficient for fitting the yield curve.[7] Hence, only a few factors can describe the yield for any maturity. Furthermore, due to the relationship between yield rates and forward rates, these factors also describes the forward curve.

Diebold and Rudebusch (2013) divides the literature of modelling the yield curve by factor models in three approaches. A first approach places structure on the estimated factors and leaving loadings free. A second approach places structure on the loadings and leaves the factors free. A third approach restrict both factors and loadings such that arbitrage possibilities are removed.

When structure is imposed on the factors, these can be given economical interpretation. The three factors in Nelson and Siegel (1987) are interpreted as a level factor, a slope factor, and a curvature factor.

The level factor of the yield curve is usually found to be non-stationary. When estimated as a fractional process, the estimate of the integration order is about 0.8, see e.g., Osterrieder and Schotman (2012). The slope and curvature factors are estimated to be non-stationary, with estimate of the fractional order between 0 and 0.5.

5.1 Theory

I start out by recalling the relationships between the yield to maturity, the forward rate and the price of a bond. Let $y(\tau)$ be the continuously compounded yield to maturity.

[6] If the co-fractional rank is set to zero and an extended parameter space is allowed for (i.e., Parameter space 2), the grid runs from 0 to 2 for both b and d. In the case where it is allowed for all the variables in Y to have different order of integration, the grid has $n + 1$ dimensions (where n is the number of variables in Y), otherwise the dimension is 2.

[7] Duffee (2011) claims that the cross sectional R^2 for factor models with a few factors are typically around 0.999.

Then, if the price of a bond is normalized to 1 when it matures, its value is given by

$$P_t(\tau) = e^{-\tau y_t(\tau)}.$$

The forward rate curve is given by

$$f_t(\tau) = \frac{P_t'(\tau)}{P_t(\tau)}.$$

Finally, it follows that the yield rate curve is given by

$$y_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(u)\mathrm{d}u.$$

Nelson and Siegel (1987) suggest the following approximation of the forward rate curve

$$f_t(\tau) = \gamma_{1t} + e^{-\lambda\tau}\gamma_{2t} + \lambda\tau e^{-\lambda\tau}\gamma_{3t},$$

which secures that all forward rates are positive. The corresponding yield curve is (see e.g. Diebold and Li , 2006)

$$y_t(\tau) = \gamma_{1t} + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\gamma_{2t} + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)\gamma_{3t} \tag{12}$$

The parameter λ represents the exponential decay rate; a small value imply slow decay and better fit at the curve at long maturities. Diebold and Li (2006) suggest fixing this at 0.7308 when τ is measured in years. The dynamic factors are represented by γ_{1t}, γ_{2t} and γ_{3t}. These are referred to as a level factor (γ_{1t}), slope factor (γ_{2t}) and curvature factor (γ_{3t}). The loading for the level factor is 1 for all values of τ. The loading for the slope factor starts at 1 and decays monotonically to 0. The loading for the curvature factor starts at 0, increases, and then decays to 0 again.

Note that $y_t(\infty) = \gamma_{1t}$. Hence, this is a long term factor and expresses the yield at long horizons. (Furthermore, $f_t(\infty) = \gamma_{1t}$, so it also express the forward rate at long horizons.) The yield and forward rate at short horizon are given by $y_t(0) = f_t(0) = \gamma_{1t} + \gamma_{2t}$. Hence, $y_t(0) - y_t(\infty) = \gamma_{2t}$, represents the slope factor. (Note that also $f_t(0) - f_t(\infty) = \gamma_{2t}$.) Finally, the butterfly spread is $2y_t(\tau^*) - y_t(0) - y_t(\infty) = \left(1 - 2e^{-\lambda\tau^*}\right)\gamma_{3t}$ for τ^* such that $2\frac{1-e^{-\lambda\tau^*}}{\lambda\tau^*} = 1$, and — hence — is proportional to the curvature factor. For $\lambda = 0.7308$ this corresponds to $\tau^* \approx 2$ and $1 - 2e^{-\lambda\tau^*} \approx 0.8$. (Similarly for the forward curve; $2f_t(\tau^{**}) - f_t(0) - f_t(\infty) = \lambda\tau^{**}\gamma_{3t}$ for τ^{**} such that $2e^{-\lambda\tau^{**}} = 1$ and for $\lambda = 0.7308$ this implies $\tau^{**} \approx 1$ and $\lambda\tau^{**} \approx 0.7$.)

Table 1: US Government Bonds

lags ℓ	rank r	likelih. L	estimates b	d	lag-reduction LR [p-val]	rank-test LR [pval]	high. ord. int. LR [p-val]
0	0	3415.8	-	0.892		136.8 [0.000]**	
	1	3479.5	0.306	0.991		9.4 [0.052]	
	2	3483.1	0.375	1.042		2.2 [0.138]	
	3	3484.2	0.463	1.163	72.6 [0.000]**		
1	0	3484.2	0.463	0.700		31.4 [0.000]**	12.519 [0.000]**
	1	3496.2	0.425	0.931		7.4 [0.116]	82.66 [0.000] **
	2	3499.8	0.476	0.966		0.2 [0.655]	107.79 [0.000]**
	3	3499.9	0.477	0.986	41.2 [0.000]**		
2	0	3499.9	0.477	0.509		22.4 [0.008]**	0.24 [0.6260]
	1	3508.3	0.443	0.443		5.6 [0.231]	0.09 [0.7611]
	2	3510.4	0.420	0.420		1.4 [0.237]	0.68 [0.4106]
	3	3511.1	0.392	0.392	18.8 [0.027]*		
3	0	3511.1	0.392	0		18.8 [0.027]*	1.49 [0.2217]
	1	3517.4	0.232	0.232		6.2 [0.185]	0.78 [0.3781]
	2	3519.6	0.306	0.306		1.8 [0.180]	1.66 [0.1978]
	3	3520.5	0.363	0.406			

The full data sample is: 1962(2) - 2013(12) (623 observations)
Sample used for initialization is: 1962(2) - 1962(12) (11 observations)
Sample used for estimation is: 1963(1) - 2013(12) (612 observations)

Table 2: US Government Bonds - co-fractional relationships

lags ℓ	rank r	co-fractional vector $\beta' Y_t =$
0	1	$y_t(3) - 1.86y_t(5) + 0.91y_t(10)$
1	1	$y_t(3) - 1.93y_t(5) + 0.92y_t(10)$
2	1	$y_t(3) - 1.22y_t(5) + 0.24y_t(10)$
3	1	$y_t(3) - 1.50y_t(5) + 0.53y_t(10)$

lags ℓ	rank r	level $\beta' = (0,0,1)$	slope $\beta' = (-1,0,1)$	curvature $\beta' = (1,-2,1)$
0	1	96.6 [0.0000]**	118.0 [0.0000]**	20.2 [0.0000]**
1	1	14.6 [0.0007]**	21.8 [0.0000]**	0.6 [0.7408]
2	1	14.0 [0.0009]**	5.2 [0.0743]	10.0 [0.0067]**
3	1	9.0 [0.0111]*	24.8 [0.0000]**	10.8 [0.0045]**

5.2 Empirical results

Here I estimate the co-fractional system with $Y_t = (y_t(3), y_t(5), y_t(10))'$, where $y_t(\tau)$ is the τ-year yield.

5.2.1 US Government Bonds

Table 1 reports the estimation results for many combinations of lag length (ℓ) and co-fractional rank (r). In estimation of the system the parameter restriction given by

Assumption 2 is imposed.

One of the identification problems — given in Corollary 2.1 — can immediately be seen from Table 1: the likelihood for the system with zero co-fractional rank is equal to the likelihood of the system with one less lag and a full co-fractional rank. The estimate of the fractional parameter b is also equal for the two systems. The estimate for the fractional parameter d differs between the two formulations, but the difference is exactly equal to the estimated b. All these results are consistent with the Proposition 2.1 and Corollary 2.1.

From the table we see that the restriction $d \geq b$ is binding for most of the systems with lag length 2 or 3 (i.e, $\ell = 2$ and $\ell = 3$) and a non-zero co-fractional rank.

In all the estimated systems reported in Table 1 the estimated b is less than 0.5. This implies that a standard χ^2-distribution can be applied for testing the co-fractional rank.

The column 'lag-reduction' in Table 1 reports lag reduction tests. The p-values are here derived from the χ^2-distribution. A lag-length of both zero or one is rejected (tested against a lag-length of 3). A lag-length of 2 can be rejected or not depending on whether a 1% or 5% significance level is applied.

The point of departure is a system with a lag length of 3. The co-fractional rank test indicates a co-fractional rank equal to one. According to Table 1 the yields are with this specification estimated to be integrated of order 0.232 and the co-fractional relationship is integrated of order 0 (as $\hat{d} - \hat{b} = 0$). The estimated co-fractional relationship for this specification is reported in Table 2. From the table we also see that we reject that this relationship corresponds to the level, trend or slope of the yield curve.

To evaluate if the results for this specification is reliable we have to test for higher order integration. This is conducted in the last column of Table 1. Here we cannot reject that $\alpha_\perp' \Gamma \beta_\perp$ has reduced rank. Hence, this indicates that our estimates of the fractional parameters d and b are not reliable (in the sense that we cannot conclude based on the estimation results that Y_t being fractionally integrated of order d and $\beta' Y_t$ being fractionally integrated of order $d - b$). Hence, to get more reliable results, the lag length is reduced to two lags.

When considering a lag length of two, the co-fractional rank test still indicate a co-fractional rank equal to one. The estimated co-fractional relationship with this specification is also reported in Table 2. For this specification we reject that the co-fractional relationship corresponds to the level or curvature. However, the test that the co-fractional relationship corresponds to the slope is not rejected for this specification. Also for this lag length higher order fractionality cannot be rejected. Hence, the lag length is reduced to 1 lag.

Table 3: Norwegian Government Bonds

lags	rank	likelih.	estimates		lag-reduction	rank-test	high. ord. int.
ℓ	r	L	b	d	LR [p-val]	LR [pval]	LR [p-val]
0	0	1723.4	-	0.925		59.9 [0.0000]**	
	1	1745.1	0.764	0.979		16.6 [0.004]**	
	2	1750.1	0.832	1.009		6.6 [0.010]*	
	3	1753.4	0.747	1.079	57.2 [0.0006]**		
1	0	1753.4	0.747	0.332		20.1 [0.0175]*	10.5 [0.001]**
	1	1759.9	0.691	0.849		7.0 [0.165]	51.5 [0.000]**
	2	1762.6	0.676	0.973		1.5 [0.197]	33.5 [0.000]**
	3	1763.4	0.785	1.063	37.1 [0.0051]**		
2	0	1763.4	0.785	0.278		25.1 [0.0029]**	2.6 [0.103]
	1	1770.7	0.894	0.894		10.5 [0.078]	48.8 [0.000]**
	2	1773.8	0.911	0.911		4.3 [0.042]*	93.3 [0.000]**
	3	1775.9	0.743	1.257	12.0 [0.2131]		
3	0	1775.9	0.743	0.514		12.0 [0.2131]	7.7 [0.005]**
	1	1779.1	0.664	0.664		5.8 [0.240]	15.3 [0.000]**
	2	1781.3	0.697	0.397		1.2 [0.250]	16.9 [0.000]**
	3	1782.0	0.709	0.709			

The full data sample is: 1987(4) - 2013(12) (321 observations)
Sample used for initialization is: 1987(4) - 1987(12) (9 observations)
Sample used for estimation is: 1988(1) - 2013(12) (312 observations))

In the specification with one lag, the co-fractional rank test also indicates a co-fractional rank of one. The estimated co-fractional relationship is close to representing the curvature, and an formal test for this is far form rejected. The test that this co-fractional relationship represents the level or slope are clearly rejected. Furthermore, with this specification the test for higher order fractionality is rejected. and and Hence, these results indicates that the curvature is integrated of a fractional order of $0.931 - 0.425 = 0.501$, whereas the level and slope are integrated of a fractional order of 0.931. This implies that the curvature component in the yield curve has a shorter memory that the level and slop components. The level and slope components are close to follow random walk processes (as they are close to being integrated of order 1), whereas the the curvature component is close to being stationary (as it wold be if it where integrated of an order less that 0.5).

5.2.2 Norwegian Government Bonds

In Table 3 and Table 4 the corresponding estimation results are reported for Norwegian government bond data. Here the data series are from a shorter period; the Norwegian data starts in the end of the 1980s whereas the US data starts in the beginning of 1960s. Hence, the US data covers about 25 more years that the Norwegian data does, such

Table 4: Norwegian Government Bonds - co-fractional relationships

lags ℓ	rank r	co-fractional vector $\beta' Y_t =$
0	1	$y_t(3) - 1.93y_t(5) + 0.93y_t(10)$
1	1	$y_t(3) - 1.89y_t(5) + 0.87y_t(10)$
2	1	$y_t(3) - 1.84y_t(5) + 0.82y_t(10)$
3	1	$y_t(3) - 1.64y_t(5) + 0.63y_t(10)$

lags ℓ	rank r	level $\beta' = (0, 0, 1)$	slope $\beta' = (-1, 0, 1)$	curvature $\beta' = (1, -2, 1)$
0	1	34.7 [0.0000]**	39.9 [0.0000]**	0.3 [0.8641]
1	1	9.7 [0.0111]**	14.0 [0.0009]**	3.0 [0.2212]
2	1	14.4 [0.0009]**	6.5 [0.0383]*	6.3 [0.0421]*
3	1	5.4 [0.0667]	2.6 [0.2791]	6.0 [0.0499]*

that the US data covers about 50 years of monthly data and the Norwegian covers about 25 years. Since the US data series has about twice as many observations as the Norwegian series, the results for the Norwegian data series may not be as reliable as those for the US data.

Table 3 can be used to examine the lag length and the co-fractional rank. A reduction from a lag length of 3 to 2 is not rejected. A lag length of 1, however, is rejected based on the χ^2-distribution, both when it is compared to a lag length of 3 or a lag length of 2. Hence, the results indicate a lang length of 2.

Given a lag length of 2, the co-fractional rank test indicates a co-fractional rank of 1 or 3. Johansen and Nielsen (2012) suggest suggest starting out by testing a rank of zero. If that rank restriction is rejected — as it is here — the next test is the test of a co-fractional rank equal to 1. This restriction is not rejected by conventional test sizes. However, if one starts out by general-to-specific approach by reducing the co-fractional rank in the null-hypothesis in each step, one ends up with a full co-fractional rank when a 5 per cent size is used in the test.

The full rank result implies that there is no co-fractional relationships between the variables. If a co-fractional rank of 1 is imposed, the estimated co-fractional relationship between the variables is given by $y_t(3) - 1.84y_t(5) + 0.82y_t(10)$, see Table 4. This is close to the butterfly spread, here used as a proxy for the curvature factor, which is $y_t(3) - 2y_t(5) + y_t(10)$. Nevertheless, this co-fractional relationship is rejected when using a test size of 5 per cent. However, when using a 1 per cent test size, neither the hypothesis that the co-fractional relationship corresponds to the slope factor nor the hypothesis that the co-fractional relationship corresponds to the curvature factor is rejected. The hypothesis that the co-fractional relationship corresponds to the level factor is clearly rejected.

Table 5: US Government Bonds, $(y_t(3), y_t(5))$

lags	rank	likelih.	estimates		lag-reduction	rank-test	high. ord. int.
ℓ	r	L	b	d	LR [p-val]	LR [pval]	LR [p-val]
0	0	2013.1	-	0.968		31.6 [0.0000]**	
	1	2020.6	0.270	1.047		16.5 [0.0000]**	
	2	2028.9	1.443	2.332	11.7 [0.1660]		
1	0	2028.9	1.443	0.888		4.3 [0.3670]	535.0 [0.000]**
	1	2030.9	0.884	0.884		0.2 [0.713]	216.3 [0.000] **
	2	2031.0	1.613	2.519	7.4 [0.1170]		
2	0	2031.0	1.613	0.906		7.4 [0.1170]	530.3 [0.000]**
	1	2034.5	0.945	0.945		0.4 [0.562]	154.8 [0.000]**
	2	2034.7	0.950	0.950			

The full data sample is: 1962(2) - 2013(12) (623 observations)
Sample used for initialization is: 1962(2) - 1962(12) (11 observations)
Sample used for estimation is: 1963(1) - 2013(12) (612 observations))

5.3 Sensitivity analysis

In a system with more than two variables there might be polynomial co-fractionality. For example, each of the variables $y_t(3)$, $y_t(5)$ and $y_t(10)$ can be integrated of order d; and any spread between them can be integrated of order $d-b$ such that $y_t(3) - y_t(10) \sim I(d-b)$ and $y_t(5) - y_t(10) \sim I(d-b)$; and finally the butterfly spread could be integrated of order $d-2b$ such that $[y_t(3) - y_t(10)] - 2[y_t(5) - y_t(10)] = y_t(3) - 2y_t(5) + y_t(10) \sim I(d-2b)$.

To investigate if there could be polynomial co-fractionality among the time series I investigate four different subsets of the variables where each subset includes only two variables. The four subsets are: $(y_t(3), y_t(5))$ in Table 5; $(y_t(3), y_t(10))$ in Table 6; $(y_t(5), y_t(10))$ in Table 7; and $(y_t(3) - y_t(10), y_t(5) - y_t(10))$ in Table 8. In all these tables I start out with a lag length of two lags.

In Table 5 co-fractionality in a system with the 3 year and 5 year bond rate is investigated. Here a reduction of the lag length to 0 is not rejected. Given this lag length, reduced co-fractional rank is rejected, indicating no co-fractional relationship among the variables. The estimated fractional order for the two yield rates are $2.332 - 1.443 = 0.888$.

In Table 6 co-fractionality between the 3 year and 10 year bond rate is investigated. Here tests for reducing the lag length from 2 are rejected, indicating a lag length of 2. Given this lag length the co-fractional rank test indicates a full rank, which implies no co-fractional relationship between the variables. The estimated fractional order of the two yield rates are $3.566 - 2.612 = 0.954$.[8]

[8]Overall, the results do not change if more lags are allowed for. With a lag length of 3, reducing the lag length to 2 is rejected. Furthermore, given a lag length of 3, a reduced co-fractional rank is rejected.

Table 6: US Government Bonds, $(y_t(3), y_t(10))$

lags ℓ	rank r	likelih. L	estimates b	d	lag-reduction LR [p-val]	rank-test LR [pval]	high. ord. int. LR [p-val]
0	0	1687.7	-	0.986		27.2 [0.0000]**	
	1	1692.0	0.736	1.010		18.6 [0.0000]**	
	2	1701.3	1.639	2.558	21.3 [0.0065]**		
1	0	1701.3	1.639	0.919		3.6 [0.470]	675.3 [0.000]**
	1	1702.7	0.914	0.914		0.8 [0.418]	257.8 [0.000] **
	2	1703.1	1.671	2.605	17.7 [0.0014]**		
2	0	1703.1	1.671	0.934		17.7 [0.0014]**	581.3 [0.000]**
	1	1706.3	0.893	1.001		11.2 [0.001]**	169.0 [0.000]**
	2	1711.9	2.612	3.566			

The full data sample is: 1962(2) - 2013(12) (623 observations)

Sample used for initialization is: 1962(2) - 1962(12) (11 observations)

Sample used for estimation is: 1963(1) - 2013(12) (612 observations))

Table 7: US Government Bonds, $(y_t(5), y_t(10))$

lags ℓ	rank r	likelih. L	estimates b	d	lag-reduction LR [p-val]	rank-test LR [pval]	high. ord. int. LR [p-val]
0	0	1916.5	-	0.933		41.6 [0.0000]**	
	1	1930.6	0.000	1.045		13.2 [0.0000]**	
	2	1937.2	1.314	2.201	13.7 [0.0907]		
1	0	1937.2	1.314	0.887		6.0 [0.1974]	291.4 [0.000]**
	1	1940.2	0.894	0.894		0.1 [0.820]	242.4 [0.000] **
	2	1940.2	0.895	0.895	7.6 [0.1054]		
2	0	1940.2	0.895	0		7.6 [0.1054]	0.1 [0.819]**
	1	1943.8	0.642	1.050		0.6 [0.428]	102.4 [0.000]**
	2	1944.1	0.582	1.177			

The full data sample is: 1962(2) - 2013(12) (623 observations)

Sample used for initialization is: 1962(2) - 1962(12) (11 observations)

Sample used for estimation is: 1963(1) - 2013(12) (612 observations))

Table 8: US Government Bonds, $(y_t(3) - y_t(10), y_t(5) - y_t(10))$

lags ℓ	rank r	likelih. L	estimates b	d	lag-reduction LR [p-val]	rank-test LR [pval]	high. ord. int. LR [p-val]
0	0	2741.8	-	0.803		94.3 [0.0000]**	
	1	2786.1	0.427	0.960		5.8 [0.0164]*	
	2	2789.0	0.61421	1.151	13.0 [0.1128]		
1	0	2789.0	0.614	0.537		4.2 [0.3743]	13.1 [0.000]**
	1	2791.1	0.575	0.575		0.0 [0.968]	14.6 [0.000] **
	2	2791.1	0.576	0.576	8.7 [0.0682]		
2	0	2791.1	0.576	0		8.7 [0.0.0682]	0.0 [0.934]
	1	2795.2	0.235	0.235		0.6 [0.4337]	2.9 [0.0889]**
	2	2795.5	0.013	0.624			

The full data sample is: 1962(2) - 2013(12) (623 observations)

Sample used for initialization is: 1962(2) - 1962(12) (11 observations)

Sample used for estimation is: 1963(1) - 2013(12) (612 observations)

In Table 7 co-fractionality between the 5 year and 10 year bond rate is investigated. Here the estimation results indicated a lag length of 0 and a full co-fractional rank. Hence, no co-fractional relationship among the variables. The estimated fractional order of integration is $2.201 - 1.314 = 0.887$.

Finally, in Table 8 co-fractionality between the difference between two yield rates are investigated. In the table this is formalized as the the 3 year and 5 year bond rate relative to the 10 year bond rate. However, it does not matter how these differences are constructed. For example, the data set $(y_t(5) - y_t(3), y_t(10) - y_t(3))$ will give exactly the same estimation results, since this data set is just a linear transformation of the data set used in the table.

The results in Table 8 indicated a lag length of 0 and a full co-fractional rank. Here we would expect a co-fractional rank equal to 1 if the butterfly spread represents a co-fractional relationship. However, if I use a 1 per cent significance level for the rank test, I do not reject a co-fractional rank equal to 1. Note that for the other subsets of the yield rates the co-fractional rank is clearly rejected. Hence, of all the four subset data set this is the only one were there are some evidence of a co-fractional relationship. When imposing one co-fractional relationship, the estimated relationship becomes $(y_t(3) - y_t(10)) - 1.7927\,(y_t(5) - y_t(10))$, which again is close to the butterfly spread. However, the hypothesis test of this parameter being equal to -2 is rejected.[9]

6 Conclusions

There are not many estimation programs that can be used to estimate co-fractional VAR models. This paper presents a new estimation program, `Cofrac`, that can be used to estimate such systems. It is menu-driven and — hence — easy to use. This is one of the contributions of this paper, and is presented in Section 4.

The program is used to identify co-fractional relationships between Government bond yield rates in US and Norway. Applying for each country a data set with bond rates with three different time to maturity, the results indicate one co-fractional relationship for both US and Norway. This relationship seems to be close to the curvature factor of the yield curve, here approximated with the butterfly spread. This another contribution of this paper, and presented in Section 5.

The estimation results illustrates that there are severe identification problems in

Finally, for this lag length, the estimated fractional order is $3.720 - 2.767 = 0.952$. Extending the lag length even further does not improve the likelihood much, hence, the null hypothesis of a lag length equal to 3 is not rejected when the alternative is 4 lags.

[9] The likelihood ratio for this test is 11.5, and with only one degree of freedom in the test, the restriction is clearly rejected.

co-fractional systems. Such identification problems are shown in Section 2. One type of identification problem is not shown previously, and is a further contribution of this paper.

To deal with the identification problems, this paper suggest to reformulate the system such that one can test for higher order integration. The absence of higher order integration secures identification in most cases for the identification problems considered here. The exception is co-fractional systems with no co-fractional relationships, which can be formulated both with a co-fractional rank of zero or a full co-fractional rank. This test for higher order integration is also a contribution of this paper, and is presented in Section 3.

References

Carlini, Federico and Paolo Santucci de Magistris (2013). On the identification of fractionally cointegrated VAR models with the $F(d)$ condition, CREATES Research Paper 2013-44, Aarhus University.

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

Diebold, F. X. and G. D. Rudebusch (2013). Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach (The Econometric and Tinbergen Institutes Lectures), Princeton University Press (January 15, 2013).

Doornik, Jurgen A. (2007). Object-Oriented Matrix Programming Using Ox, 3rd ed., Timberlake Consultants Press, London, England.

Duffee, G. R. (2011). Forecasting with the term structure: the role of no-arbitrage restrictions, Working paper, the Johns Hopkins University, Department of Economics, No. 576.

Giese, Julia V. (2008). Level, Slope and Curvature: Characterising the Yield Curve in a Cointegrated VAR Model. Economics: The Open-Access, Open-Assessment E-Journal 2, number 2008-28.

Jensen, A.N. and M.. Nielsen (2014). A fast fractional difference algorithm. Journal of Time Series Analysis 35, 428-436.

Johansen, Søren (1988). Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control, 12:231-254.

Johansen, Søren (1996). Likelihood-based inference in cointegrated vector autoregressive models, Oxford: Oxfort University Press.

Johansen, Søren (2008). A representation theory for a class of vector autoregressive models for fractional processes, Econometric Theory, 28:651-676.

Johansen, Søren and Morten Ørregaard Nielsen (2010). Likelihood inference for a nonstationary fractional autoregressive model, Journal of Econometrics 158:51-66.

Johansen, Søren and Morten Ørregaard Nielsen (2012). Likelihood inference for a fractional cointegrated vector autoregressive model, Econometrica 80:2667-2732.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curve. Journal of Business 60, pp. 473-489.

Nocedal, J. and S. J. Wright (1999). Numerical Optimalization. New York: Springer-Verlag.

Osterrieder, D. and P. C. Schotman (2012). The volatility of Long-term Bond Returns: Persistent Interest Shocks and Time-varying Risk Returns, CREATES Research Paper 2012-35.

Tschernig, Rolf, Enzo Weber, Roland and Weigand (2013). Long-Run Identification in a Fractionally Integrated System, Journal of Business amp; Economic Statistics, 31(4):438-450

A Proofs of propositions

The formulation of the fractionally cointegrated VAR in (1) includes both a fractional difference operator and a fractional lag operator. It will be convenient to also have a formulation of the co-fractional system where only the fractional difference operator is used. By inserting for the expression for the lag operator and reformulate, we get

$$\sum_{i=-1}^{\ell} \Psi_i \Delta^{d+ib} Y_t = \varepsilon_t, \tag{13}$$

where $\Psi_{-1} = -\Pi = -\alpha\beta'$, $\Psi_0 = I_n - \sum_{i=1}^{\ell} \Gamma_i + \Pi = -\Gamma + \Pi$ and $\Psi_\ell = (-1)^{\ell+1}\Gamma_\ell$ defines some of the parameter matrices in (13). Furthermore, we have $\sum_{i=0}^{\ell} \Psi_i = I_n$. Now, define $\lambda_*^{(1)} = \lambda_*^{(1)}(\ell^{(1)}, r^{(1)}) = (d^{(1)}, b^{(1)}, \Psi_{-1}^{(1)}, \ldots, \Psi_{\ell^{(1)}}^{(1)})$ be parameter set 1 with

this formulation of the system.[10] The proofs of the propositions are now derived with these transformed parameter sets, where parameter sets 2 and 3 are defined similarly as parameter set 1.

A.1 Proof of Proposition 2.1

Proof. With the transformed parameter set, parameter set 1 is given by $\lambda_*^{(1)}$. We shall show that parameter set 2, given as $\lambda_*^{(2)}$, yields an identical description of the system with $d^{(2)} = d^1 - kb^{(1)}$, $b^{(2)} = b^{(1)}$ and $\ell^{(2)} = \ell^{(1)} + k$, where k is a positive integer.

With transformed parameter set $j, j = 1, 2$, (13) can be formulated as

$$\sum_{i=-1}^{\ell_j} \Psi_i^{(j)} \Delta^{d^{(j)}+ib^{(j)}} Y_t = \varepsilon_t$$

Applying parameter set 2 in (13) together with $d^{(2)} = d^{(1)} - kb^{(1)}$, $b^{(2)} = b^{(1)}$ and $\ell^{(2)} = \ell^{(1)} + k$:

$$\begin{aligned} \sum_{i=-1}^{\ell^{(2)}} \Psi_i^{(2)} \Delta^{d^{(2)}+ib^{(2)}} Y_t &= \sum_{i=-1}^{\ell^{(1)}+k} \Psi_i^{(2)} \Delta^{d^{(1)}+(i-k)b^{(1)}} Y_t \\ &= \sum_{i=-1-k}^{\ell^{(1)}} \Psi_{i+k}^{(2)} \Delta^{d^{(1)}+ib^{(1)}} Y_t \end{aligned}$$

Comparing this with the similar formulation with parameter set 1 yields:

$$\begin{aligned} \sum_{i=-1}^{\ell^{(1)}} \left(\Psi_i^{(1)} - \Psi_{i+k}^{(2)} \right) \Delta^{d^{(1)}+ib^{(1)}} Y_t &= \sum_{i=-k-1}^{-2} \Psi_{i+k}^{(2)} \Delta^{d^{(1)}+ib^{(1)}} Y_t \\ &= \sum_{i=-1}^{k-2} \Psi_i \Delta^{d^{(1)}+(i-k)b^{(1)}} Y_t \end{aligned}$$

From the last expression we see that the equality holds if $\Psi_{i+k}^{(2)} = \Psi_i^{(1)}$ for $i = -1, \ldots, \ell^{(1)}$, and $\Psi_j^{(2)} = 0$ otherwise (i.e., for $j = -1, \ldots, k-2$).

For $k = 1$ the parameter mapping implies $\Psi_0^{(2)} = \Psi_{-1}^{(1)}$. The mapping between the transformed and non-transformed parameter sets involves $\Psi_{-1}^{(1)} = \Pi^{(1)}$ and $\Psi_0^{(2)} = \Pi^{(2)} - \Gamma^{(2)}$. Finally, we have $\Psi_{-1}^{(2)} = \Pi^{(2)} = 0$. Combining these yields $\Gamma^{(2)} = \Pi^{(2)}$.

For $k > 1$ the parameter mapping implies $\Psi_0^{(2)} = 0$. Otherwise, the relationships between the parameters are as for $k = 1$. Combining these yields $\Gamma^{(2)} = 0$.

[10] We could have excluded one of the Ψ's in this parameter set due to the identity $\sum_{i=-1}^{ell} \Psi_i = I_n$.

Since the parameter mappings implies that the right hand side of (13) are identical for the two parameter spaces, the errors on the left hand side of (13) must be identical as well. This implies $\Omega^{(2)} = \Omega^{(1)}$. ■

Remark A.1.1 *The parameter space for d given by Parameter space 1 or 2 will imply a limit to how many such equivalent formulations that exists.*

A.2 Proof of Proposition 2.2

Proof. We shall prove that parameter set 3, $\lambda_*^{(3)}$, gives an identical system as parameter set 1, $\lambda_*^{(1)}$, when $d^{(3)} = d^{(1)} - \frac{k-1}{k}b^{(1)}$, $b^{(3)} = b^{(1)}/k$ and $\ell^{(3)} = (k+1)\ell^{(1)} - 1$.

Applying $d^{(3)} = d^{(1)} - \frac{k-1}{k}b^{(1)}$, $b^{(3)} = b^{(1)}/k$ and $\ell^{(3)} = (k+1)\ell^{(1)} - 1$ for parameter set 3 in (13) yields:

$$\sum_{i=-1}^{\ell^{(3)}} \Psi_i^{(3)} \Delta^{d^{(3)}+ib^{(3)}} Y_t = \sum_{i=-1}^{(k+1)\ell^{(1)}-1} \Psi_i^{(3)} \Delta^{d^{(1)}+\frac{i-k+1}{k}b^{(1)}} Y_t$$

Setting this equal (13) for parameter set 1 yield

$$\sum_{i=-1}^{\ell_1} \left(\Psi_i^{(1)} - \Psi_{k(i+1)-1}^{(3)} \right) \Delta^{d^{(1)}+ib^{(1)}} Y_t \;\; = \sum_{\substack{i=-1 \\ i \neq k-1, 2k-1, \ldots, \ell_1 k-1)}}^{(k+1)\ell_1-2} \Psi_i^{(3)} \Delta^{d^{(3)}+ib^{(3)}} Y_t$$

From the last expression we see that the equality holds if $\Psi_{5,k(i+1)-1} = \Psi_{1,i}$ for $i = 1, \ldots, \ell_1$, and $\Psi_{5,i} = 0$ otherwise.

Note that $\Psi_{-1}^{(3)} = -\Gamma^{(3)}$ and $\Psi_0^{(3)} = 0$ for any applicable k (i.e., an integer exceeding one). Furthermore, since $\Psi_{-1}^{(1)} = -\Gamma^{(1)}$ and $\Psi_{-1}^{(1)} = -\Psi_{-1}^{(3)}$ combined with $\Psi_0^{(3)} = -\Gamma^{(3)} + \Pi^{(3)}$, we have $\Gamma^{(3)} = -\Pi^{(1)}$. ■