A Fractional Cointegration Analysis of the Term Structure of Interest Rates.

Fabrizio Iacone*
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE

November 2, 2004

Abstract

The short end of the US term structure of interest rates over the period 1979-2002 is analysed allowing the possibility of fractional integration and cointegration of the data. The application of the fractional model is of particular importance in this case because it helps to reconcile two apparently conflicting groups of theories about the order of integration of the interest rates. Since in fact the central bank manages the short term interest rate to stabilise the inflation with a (supposedly) linear rule, we expect for those two time series a mean reverting behaviour. Dynamic models of the term structure of interest rates, on the other hand, are formulated with a cointegrated form, so if we are restricted to $I(0)$ and $I(1)$ processes only the two approaches are incompatible.

We estimated the order of fractional integration of the data, finding it compatible with a mean reverting model; we also obtained evidence in favour of cointegration. The restriction that the data were $I(1)$ and the residuals were $I(0)$ was rejected, mainly because the latter still displayed long memory.

Keywords: Long memory, Fractional Cointegration, term structure.

JEL classification: C22.

1 Introduction

There is a widespread consensus on the prescription that the monetary authority should aim at inflation stabilization in the medium term. In the long run, then, a tendency for inflation to revert to a central tendency should emerge, and it is often argued that such

---

*I am very grateful to P.M. Robinson for many helpful discussions and comments. This research was supported by ESRC Grant R000239936 and the Denis Sargan Memorial Fund. Tel: +44-20-7955-7279, fax: +44-20-7831-1840/ E-mail: f.iacone@lse.ac.uk
a characteristic should emerge for the interest rate too, due to the usual assumption of some type of linear relation between the two variables. When only $I(1)$ and $I(0)$ representations are allowed for, mean reversion is only possible with the second model, which may then be preferred despite the Dickey and Fuller test usually does not reject the hypothesis of unit root: Clarida, Gali and Gertler (2000), for example, adopted the $I(0)$ representation arguing that the results of the test were due to the low power rather than to an effective unitary root.

The interest rate considered by Clarida et al. is the Federal Funds rate, which is defined for interbank deposits with very short maturity. The choice is sensible, because the central bank actually operates supplying liquidity on the interbank market with open market operations and discount window loans, so this rate is directly controlled by the monetary authority. Yet the relevant macroeconomic indicators are only affected by the rates of contracts having much longer maturities: the demand of money from the individual agents depends on the return of the alternative asset, which is more likely to be the Treasury bills or Treasury bonds rate, while the economic activity depends rather on the bank loan rate or on the commercial paper rate. The transmission of monetary policy then requires that an impulse originated in a market for a very short term contract is transmitted to the ones having longer maturity, and for the policy to be successful it is necessary that the transmission is fast and reliable, so the disturbances to the relations linking the rates of contracts with short and long maturities should not remain in the system for much time. Plotting together two comparable interest rates with different maturities, it is indeed apparent that the long term dynamics of the two series is largely driven by an highly persistent common factor, while the persistence of the difference (the spread) is much lower. This is clearly an intuitive definition of the concept of cointegration, and it is not surprising then that an application for two rates with different maturity is already present in the seminal paper of Engle and Granger (1987).

Campbell and Shiller (1987) provided a theoretical model linking rates with different maturities: they showed that introducing rational explanations and allowing for a constant or $I(0)$ risk premium, the interest rate spread has $I(0)$ structure even when the original data are $I(1)$, in which case $C(1,1)$ cointegration results.

We review the references to the monetary theory in greater details in Section 2, describing alternative models for the interest rates and paying particular attention the empirical analyses dedicated to the dynamic transmission of impulses across the different maturities.

Some of the implications of the monetary policy prescriptions are at odds with the empirical analyses, but this only depends on the excessive restriction implicit in that framework. In Section 3 we show that allowing for fractional integration and cointegration is a way to solve this apparent conflict, combining mean reversion with persistence of the innovations and common long term dynamics. We review the definition of integration and cointegration presenting a formulation that is general enough to encompass not only the $I(0)$ and $I(1)$ structures, but also other $I(d)$ ones having non integer $d$. We then describe some semiparametric and parametric techniques to estimate the order of integration, the cointegration rank and the parameters corresponding to the long term restrictions, and finally the whole parametric model.
In Section 4 we present the application the fractional cointegration model to the short end of the US term structure. We estimated the order of fractional integration of the data, finding it compatible with a mean reverting model; we also obtained evidence in favour of cointegration. The restriction that the data were \( I(1) \) and the residuals were \( I(0) \) was rejected, mainly because the latter still displayed long memory. We thus found the data incompatible with the expectations hypothesis, but the link between rates with different maturities appeared strong enough to allow a reliable and fast transmission of the monetary impulses.

2 Dynamic Modelling of Interest Rates in the Applied Literature

It is generally acknowledged that one of the tasks of monetary policy is the stabilization of inflation over time. Inflation should fluctuate around the target, and in the medium run any overshooting or undershooting should disappear.

It is possible to argue that such a mean reverting behaviour should pass to the interest rates too. This is the case, for example, if the Fisher equation holds: let \( i_t \) be the interest rate over the period starting in \( t \), \( \pi_{t+1} \) the one-period inflation observed in \( t+1 \), and \( E_t(\pi_{t+1}) \) the expectation formulated with the information available at time \( t \); then

\[
i_t = E_t(\pi_{t+1}) + r
\]

where \( r \) is defined as the real interest rate. A milder version of the Fisher equation allows some short run deviations \( \eta_t \):

\[
i_t = E_t(\pi_{t+1}) + r + \eta_t.
\]

A further generalization is due to Taylor (1993), who proposed the model

\[
i_t = \mu_t + \nu_\pi \pi_t + \nu_y y_t + \eta_t
\]

where \( y_t \) is a vector of other variables, such as the state of the economy, and the impact of inflation on the interest rate is not necessarily one. Svensson (1997) provided a theoretical framework in which such a policy emerges as the optimal behaviour for a central bank that wants to stabilize inflation, and Clarida et al. considered an alternative version in which \( \pi_t \) is replaced with its expectation.

In all these models the order of integration of the interest rate is indirectly providing a piece of information about the one of the inflation: if the target does not change over time and the central bank is successful in its action, the order of integration of the interest rate should be below one, allowing for mean reversion at least in the long run. If we rule out the fractional model, we are then forced to model the interest rate as an \( I(0) \).

When vectors of interest rates are considered, it is important to model the interactions between them too. Letting \( P_{t,t+k} \) be the price of a bond that is bought at the time \( t \) and
that allows the owner a claim of 1 at time \( t + k \), the continuously compounded interest rate associated to this bond is

\[
i_{t; t+k} = -\frac{1}{k} \ln P_{t; t+k}.
\]

If an economic agent wants to invest a certain amount for the period \( t, t + k \), she can choose between buying one single contract now, gaining \( i_{t; t+k} \) over the each period until the end, or buying one-periods contracts each time, gaining \( i_{t; t+1}, i_{t+1; t+2} \). Since \( i_{t+1; t+2} \) can not be observed at time \( t \), the economic agent has to consider \( E_t(i_{t+1; t+2}) \), the rate expected at time \( t \) for the period \( t + 1; t + 2 \). The two strategies then only give the same expected profit if

\[
i_{t; t+2} = \frac{1}{2} (i_{t; t+1} + E_t(i_{t+1; t+2})).
\] (1)

The equation (1) is often augmented with a drift term \( lp \)

\[
i_{t; t+2} = \frac{1}{2} (i_{t; t+1} + E_t(i_{t+1; t+2})) + lp
\] (2)

to allow for a liquidity premium, assuming a preference for liquidity of the economic agents. The liquidity premium is also relevant theoretically, as it may follow from the concavity of the utility function. The model is completed assuming rational expectations:

\[
E_t(i_{t+1; t+2}) = i_{t+1; t+2} - \eta_{1,t+1}
\] (3)

where \( \eta_{1,t} \) is an i.i.d. process; in a weaker version of the expectations model, (2) is also subject to i.i.d. disturbances \( \eta_{2,t} \).

The expectations theory then does not impose any specific restriction on the order of integration of the data. If anyway the data are integrated, then the model implicitly requires cointegration: substituting (3) in (2) and adding and subtracting \( \frac{1}{2} i_{t; t+1} \), we get

\[
i_{t; t+2} = lp + \nu i_{t; t+1} + \eta_t
\] (4)

where \( \nu = 1 \) and \( \eta_t = \eta_{2,t} - \frac{1}{2} \Delta i_{t+1; t+2} - \eta_{1,t+1} \), so assuming an \( I(d) \) structure for the interest rate, the residuals are then of order \( I(max \{d - 1, 0\}) \). The model can of course also be specified for contracts with longer maturities, \( i_{t; t+k} \), and Campbell and Shiller show how this translates in the equations (1) - (4).

Many tests were proposed for the expectation theory hypothesis.

Under integration, \( i_{t; t+k} \) and \( i_{t; t+1} \) should be cointegrated with long term coefficient \( (1, -1) \), and this was often presented as a preliminary test. When only pairs of interest rates are considered, the evidence in favour of the hypothesis is widespread, and appears to be quite robust with respect to changes of the pairs of rates used (Treasury bills and bonds, eurodeposits, synthetic rates generated fitting splines) and to the sample period. Increasing the dimension of the vector of interest rates, the cointegration rank should follow, leaving only one single common stochastic trend to drive the system: Hall, Anderson and Granger (1992), Engsted and Tangaard (1994) and Lanne (2000) found
more than one stochastic trend, but explained the result allowing for a structural break in 1979 (Hall et al. introduced a break in 1982 too); Domingues and Novales (2000) on the other hand found no stochastic trend at all when the euro-deposit rate was considered and the sample started in 1979, but they interpreted it as evidence in favour of one stochastic trend rather than for an I(0) structure, having taken into account the Dickey and Fuller tests (they also noticed that, starting the sample just a few years later, the cointegration test confirmed the hypothesis of one single stochastic trend). The potential breaks were explicitly estimated by Hansen (2003), who found that the two changes in monetary policy in 1979 and in 1982 altered the short run dynamics of the rates, while leaving the long run characteristics (one single stochastic trend, (1,−1) cointegrating vectors) unchanged.

Although most of the researchers assumed an I(1)/I(0) structure and adopted the pseudo - maximum likelihood approach discussed by Johansen (1991), alternative techniques were occasionally introduced to model some features: arguing that the I(1) representation contrasts the experience that rates do not take negative values, Lanne allowed for near unit roots and introduced a test robust with respect to deviations from the exact unit root assumption.

A second group of empirical works aimed testing the expectation hypothesis analysing the implication of that theory on the short run dynamics of the data.

Cambell and Shiller considered a VAR for the spread \( (i_{t:t+k} - i_{t:t+1}) \) and the interest rate variation \( (i_{t:t+1} - i_{t-1:t}) \), showing that the expectations theory imposes a particular set of restrictions to the coefficients of the auto regressive structure, and finding that those restrictions corresponding were rejected by the data. In a single equation framework, Campbell and Shiller (1991) and Campbell (1995) considered two very popular tests, focussed on the ability of the spread to predict the long term interest rate variation \( (i_{t+1:t+k} - i_{t:t+k}) \) and the variation of the short term rate in the long run \( \sum_{j=1}^{k} (i_{t+j:t+j+1}/k - i_{t:t+1}) \); in both cases they found that the reaction is lower, or that it even goes in the wrong direction; Jondeau and Ricart (1999) took the future rates, that are observable on the market, explicitly into account, but they too rejected the expectations hyptotesis. Using a different setting, Longstaff found that the expectation error \( \sum_{j=1}^{k} (i_{t+j:t+j+1}/k - i_{t:t+k}) \) had zero mean and was not correlated to \( i_{t:t+k} \) itself, thus providing support for the theory.

Summarising the empirical analyses we surveyed, the CI(1, 1) design seems to prevail when the expectations hypothesis and the transmission mechanism are discussed jointly. Although the results seem to have a certain sensitivity the contracts considered and to the time spanned by the sample period, the evidence of a long run relation linking the rates across the term structure is quite convincing, albeit the expectation theory is usually rejected when the short term relations between rates are taken into account. As Campbell and Shiller pointed out, anyway, the failure to observe some strong implications of the expectations theory does not necessarily mean that it plays no role at all in the explanation of the interest rate dynamics.
3 Fractional Integration and Cointegration

We consider a scalar, covariance stationary process \( \eta_t \) with mean zero and spectral density \( f_{\eta\eta}(\lambda) \) defined as

\[
E(\eta_t \eta_{t+j}) = \int_{-\pi}^{\pi} e^{ij\lambda} f_{\eta\eta}(\lambda) d\lambda
\]

and we say that the process \( \eta_t \) is integrated of order 0 if \( f_{\eta\eta}(\lambda) \) is finite and continuous and nonzero at all the frequencies \( \lambda \). Since in general we use the notation \( I(d) \) to indicate that a process is integrated of order \( d \), \( \eta_t \) is \( I(0) \). We introduce the indicator function \( 1(\cdot) \) and the difference operator \( \Delta = 1 - L \), where \( L \) is the lag operator \( (L\eta_t = \eta_{t-1}) \): an \( I(d) \) process is then defined for \( d > 0 \) for a scalar time series \( \xi_t \) as

\[
\xi_t = \mu_\xi + \Delta^{-d} \{\eta_1(t > 0)\}, \quad d > 0, \quad t \in \{0, \pm 1, \pm 2, \ldots\},
\]

where \( \mu_\xi \) is a generic constant; when \( d \) is not integer, the expansion

\[
(1 - z)^d = \frac{1}{\Gamma((-d))} \sum_{j=0}^{\infty} \frac{\Gamma(j - d)z^j}{\Gamma(j + 1)} \quad \text{where} \quad \Gamma(\alpha) = \int_0^\infty e^{-s}s^{\alpha-1}ds
\]

is considered.

Strictly speaking the truncation of the values before \( t = 1 \) makes the process \( \xi_t \) non-stationary even when \( d < 1/2 \): a definition of a stationary fractional process is possible and is in fact widely considered in the literature, and it can be also extended to deal with \( d \geq 1/2 \) too, as discussed by Velasco (1999a), but we present (2) because it does not impose the to distinguish between \( d < 1/2 \) and \( d \geq 1/2 \).

Due to the truncation, the process in (6) does not have a constant spectral density, but Robinson and Marinucci (2001) considered a “time varying” one, \( f_{\xi\xi}^{(T)} \) (where \( T \) is the dimension of the sample size), and showed that in general the properties of the processes defined in the stationary and in the truncated way (when \( 0 < d < 1/2 \)) are equivalent, and in particular, for a certain positive constant \( G_{\xi\xi} \),

\[
f_{\xi\xi}^{(T)}(\lambda) \sim G_{\xi\xi}\lambda^{-2d} \quad \text{when} \quad \lambda + (\lambda T)^{-1} \to 0^+,
\]

where \( \sim \) means that the ratio at the left and at the right hand side tend to 1 under the given condition. Since in the empirical analysis we also deal with first differences of fractional processes, we remark that the definition in (2) can be extended to \( -1/2 < d < 0 \).

Values of \( d \) between 0 and 1 define processes having properties intermediate between the \( I(0) \) and the \( I(1) \) ones: the spectral density (or its time varying version, or the pseudospectrum) gets steeper when \( d \) increases in a neighborhood of \( \lambda = 0 \); the process remains mean reverting as long as \( d < 1 \), but the speed with which the reversion takes place decreases with the increase of \( d \), and the properties of the sample mean and of the
sample variance depend on $d$ too; extensive and very detailed discussions are in Taqqu (1975) and in Hosking (1996) for the stationary case, in Velasco (1999a) for the extension to larger $d$, and Marinucci and Robinson (2000) for the truncated process.

A multivariate process $x_t$ can be defined simply stacking $n$ scalar $x_{1,t}, \ldots, x_{n,t}$ ones: $x_t = (x_{1,t}, \ldots, x_{n,t})'$. Assuming that $x_{1,t} \in I(d_1), \ldots, x_{n,t} \in I(d_n)$, then $x_t \in I(d_1, \ldots, d_n)$ and (8) can be rewritten as $f_{xx}^{(T)}(\lambda) \sim \Delta(\lambda) G_{xx} \Delta^*(\lambda)$ when $\lambda + (\lambda T)^{-1} \to 0^+$, where $G_{xx}$ is positive semidefinite with nonzero elements on the main diagonal, $\Delta(\lambda) = \text{diag} \{ e^{i\pi d_1/2} \lambda^{-d_1}, \ldots, e^{i\pi d_n/2} \lambda^{-d_n} \}$ and $\Delta^*(\lambda)$ is its complex conjugate.

When in particular
\[ d_1 = \ldots = d_n = d \]
and there is a $n \times r$ matrix $\beta$ of rank $r$, $0 < r < n$, and a linerly independent $n \times (n - r)$ matrix $\alpha$ of full rank, such that
\[ \alpha' x_t \in I(d, \ldots, d) \text{ and } \beta' x_t \in I(b_1, \ldots, b_r) \text{ with } 0 \leq b_1 \leq b_2 \ldots < d, \]
then at least some of the elements of the vector $x_t$ are cointegrated, and we indicate this as $CI(d, \ldots, d - b_1, \ldots, d - b_r)$ (this is essentially based on the definition of Breitung and Hassler, a more general definition of cointegration was provided by Robinson and Yajima, 2002).

In this case we assume that the elements of $x_t$ are ordered as
\[ x_t = (y_t', z_t')', \]
where $y_t$ is a $r \times 1$ vector and $z_t$ is $(n - r) \times 1$, and $G_{zz}$ is full rank (so that the elements in $z_t$ are not cointegrated), then considering the model
\[
\begin{align*}
y_t &= \nu' z_t + \mu_y + \Delta(-b_1, \ldots, -b_r) u_{y,t} \\
z_t &= \mu_z + \Delta(-d, \ldots, -d) u_{z,t}
\end{align*}
\]
where $\mu_y$ and $\mu_z$ are two vectors of constants, $\Delta(-b_1, \ldots, -b_r)$ is an $r \times r$ matrix having $\text{diag}\{\Delta^{-b_1}, \ldots, \Delta^{-b_r}\}$ and zeros otherwise, and similarly $\Delta(-d, \ldots, -d) = \text{diag}\{\Delta^{-d}, \ldots, \Delta^{-d}\}$ of dimension $(n - r) \times (n - r)$, and
\[ u_t = (u_{y,t}', u_{z,t}')' \]
is an I(0) process. Also notice that although different $b_1, \ldots, b_r$ are possible, in the rest of the section we assume $b_1 = \ldots = b_r = b$ to shorten the notation.

Both parametric and semiparametric techniques were developed to estimate memory parameters or the cointegrating vectors and the number of cointegrating restrictions. The semiparametric techniques are less efficient, sometimes even resulting in a lower rate of convergence, but have the advantage to focus on the characteristics of interest only, being then robust and faster in the application. The fully parametric approach on the other hand requires the specification of the whole structure, forcing the researcher to model also features of no importance for her analysis; this is clearly a nuisance, because it may require some time, but it has the additional drawback that a misspecification of
the characteristics to which the researcher is not interested may cause the inconsistency
in the estimation of the parameters of interest.

Consistent and asymptotically normal distribution of the estimate of the memory
parameter was proposed by Fox and Taqqu (1986) for a parametric model for stationary
and scalar gaussian process; Giraitis and Surgalis (1991) relaxed the assumption of
gaussianity, and Velasco and Robinson (2000) extended the results to non-stationary
data; multivariate processes can be dealt with in the same way as long as the rank of $G_{xx}$
is full. Semiparametric techniques were then developed too, focussing on the parameters
of interest only: Geweke and Porter-Hudak (1983) and Robinson (1995a) discussed the
log-periodogram regression estimate of the memory parameter, while Robinson (1995b)
proposed the local Whittle one, which is characterised by a lower variance; a multivariate
version of this estimator was later discussed by Lobato (1999).

Least Squares estimation of the cointegrating vector was already proposed by Engle
and Granger, and, assuming prior knowledge of $d$, $b$, a semiparametric test for coin-
tegration based on the estimation of the rank of the matrix $G_{xx}$ was introduced by
Phillips and Ouliaris (1988); a fully parametric approach, aimed at simultaneously esti-
mation of the cointegration rank and vector when $d = 1$, $b = 0$ are known, is in
Johansen (1991). Allowing for fractional processes, a semiparametric estimator of the
cointegrating parameter was proposed by Robinson (1994) and discussed by Robinson
and Marinucci (2001), while semiparametric estimation of the cointegration rank was
discussed by Robinson and Yajima (2002); a two step, fully parametric estimator of
the cointegrating vector was discussed by Robinson and Hualde (2003), who presented
a set of sufficient conditions to make it as efficient as the one estimated by maximum
likelihood, while a parametric test for cointegration is in Breitung and Hassler (2002).

In order to test for cointegration, we first have to estimate the memory of the
processes involved. Several techniques were proposed in the literature to address this is-
swue: here we consider the local Whittle estimate $d$, derived in Robinson (1995b), because
it has the lowest asymptotic variance.

To formally describe the technique, we consider a generic process $x_t$ observed at the
times $t \in \{1, 2, ...T\}$: the Fourier transform computed for the frequency $\lambda$ is then defined as

$$w_x(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^{T} x_t e^{i\lambda t}$$

and the periodogram is

$$I_{xx}(\lambda) = w_x(\lambda)w_x(-\lambda)^\prime.$$

Robinson considered the Whittle approximation of the gaussian log-likelihood on the
lowest $m$ frequencies, and Lobato (1999) discussed the generalisation to a multivariate
vector, proposing the loss function

$$L(\theta_{d1}, ..., \theta_{dn}, \theta_G) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \ln |\Lambda(\lambda_j) G\Lambda(\lambda_j)| + tr \left[(\Lambda(\lambda_j) G\Lambda(\lambda_j))^{-1} I_{xx}(\lambda_j)\right] \right\}$$
where $G$ is assumed positive definite, and

$$
\Lambda(\lambda_j) = diag(\lambda_j^{-\theta_{d1}}, ..., \lambda_j^{-\theta_{dn}}).
$$

Concentrating

$$
\hat{G}(\theta_{d1}, ..., \theta_{dn}) = \frac{1}{m} \sum_{j=1}^{m} \text{Re}(\Lambda(\lambda_j)^{-1} I_{xx}(\lambda_j)(\Lambda(\lambda_j)^{-1}))
$$

we define

$$
(\vec{d}_2, ..., \vec{d}_n) = \arg \min_{\theta_{d1}, ..., \theta_{dn}} \left\{ \ln |\hat{G}(\theta_{d1}, ..., \theta_{dn})| - \frac{1}{m}(\theta_{d1}, ..., \theta_{dn}) \sum_{j=1}^{m} \ln(\lambda_j) \right\}.
$$

(15)

Under the condition that $m \to \infty$ and $\frac{m}{T} \to 0$ when $T \to \infty$, $-1/2 < d_1 < 1/2$, ..., $-1/2 < d_n < 1/2$,

$$
\sqrt{m}(d_1, ..., d_n)' - (d_1, ..., d_n)' \to_d N(0, E^{-1})
$$

where $E = 2(I_p + G_{xx} \circ G_{xx}^{-1})$ and $\circ$ denotes the Hadamart product of the two matrices. Two cases of particular interest are $n = 1$ and $n = 2$: in the first case, $E^{-1} = \frac{1}{4}$, while if $n = 2$

$$
E^{-1} = \frac{1}{8} \begin{bmatrix}
2 - c^2 & c^2 \\
c^2 & 2 - c^2
\end{bmatrix} \text{ where } c^2 = \frac{G_{1,2}^2}{G_{1,1}G_{2,2}}
$$

and $G_{k,l}$ are the elements in the position $k, l$ of the matrix $G_{xx}$.

Since the loss function is not linear in $\theta_{d1}, ..., \theta_{dn}$, no closed form solution exists, and we initialase the numeric optimization procedure with the log-periodogram regression estimate $\hat{d}$.

Consistency and limit normality can be extended if the mean of the first difference is 0 (that is, the data do not have a linear time trend), as discussed in Velasco (1999a) and (1999b): consistency follows for $d < 1$ ($d \leq 1$ for the log-periodogram regression) while the asymptotic normality of the distribution holds if $d < 0.75$. Velasco also showed how to extend the set of potential $d$ to higher values: introducing the tapered Fourier transform

$$
w_{x, T_a}^T(\lambda_j) = \frac{1}{\sqrt{2\pi \sum_{t=1}^{T} h_t^2}} \sum_{t=1}^{T} h_t x_t e^{i \lambda_j t}
$$

and the periodogram $I_{xx}^T(\lambda_j) = w_{x, T_a}^T(\lambda_j)w_{x, T_a}^T(-\lambda_j)'$, he proposed to estimate $d_1, ..., d_n$ with log-periodogram regression (1999a) or with local Whittle (1999b) using only the frequencies $j = p, 2p, ..., m$, where the integer $p$ is the order of the taper (particular tapers are necessary to remove deterministic trends of order $t, t^2, ...$ if they are present). Tapering anyway causes an increase of the variance of a factor $p \Phi$ ($\Phi$ being specific to each taper), so it should only be considered a preliminary step to assess a range
for $d$, after which the standard local Whittle procedure can be confidently applied. A particular taper is the cosine bell,

$$h_t = \frac{1}{2}(1 - 2\cos\frac{2\pi t}{T}),$$

for which $p = 3$, $\Phi = 1$ or $p = 1$, $\Phi = 32/15$.

To test the condition (9) Robinson and Yajima considered the estimates $\widehat{d}_1, \ldots, \widehat{d}_n$, generated with the univariate local Whittle approach (notice that the multivariate local Whittle is not bound to give consistent estimates of $d_1, \ldots, d_n$ because $G_{xx}$ does not have full rank). For the simple hypothesis

$$H_0 : \{d_k = d_l\}$$

they proposed

$$\widetilde{T}_{k,l} = \frac{m^{1/2}(|\widehat{d}_k - \widehat{d}_l|)}{\left\{\frac{1}{2}(1 - \widehat{G}_{k,l}^2/(\widehat{G}_{k,k} \widehat{G}_{l,l}))\right\}^{1/2}}$$

and showed that, given regularity conditions,

$$\widetilde{T}_{k,l} \rightarrow_d N(0, 1)$$

when $x_{k,t}$ and $x_{l,t}$ are not cointegrated.

Since the distribution of $\widetilde{T}_{k,l}$ is not well defined under cointegration, they also suggested

$$\widehat{T}_{k,l} = \frac{m^{1/2}(|\widehat{d}_k - \widehat{d}_l|)}{\left\{\frac{1}{2}(1 - \widehat{G}_{k,l}^2/(\widehat{G}_{k,k} \widehat{G}_{l,l}))\right\}^{1/2} + h(T)},$$

where $h(T)$ is a sequence going to zero with an appropriate speed, because

$$\widehat{T}_{k,l} \rightarrow_p 0 \text{ if } x_{k,t}, x_{l,t} \text{ are cointegrated}$$

and

$$\widehat{T}_{k,l} \rightarrow_d N(0, 1) \text{ if } x_{k,t}, x_{l,t} \text{ are not cointegrated but } d_k = d_l.$$ 

They remarked anyway that $|\widehat{T}_{k,l}| < |\widetilde{T}_{k,l}|$ so that a non rejection of the hypothesis in (9) from $\widetilde{T}_{k,l}$ would be made with even greater confidence from $\widehat{T}_{k,l}$.

When $n$ variables are considered, it can be of interest to jointly test up to $n - 1$ hypothesis: they then extended (16) as

$$\widehat{T} = (S \widehat{d})' (S \widehat{D}^{-1}(\widehat{G}_{xx} \circ \widehat{G}_{xx})\widehat{D}^{-1} S'/(4m) + h^2(T)I_{n-1})^{-1} (S \widehat{d})$$

where $\widehat{d}$ is the vector where the estimated memory parameters are stacked, $S$ is an $(n - 1) \times n$ matrix with $S_{k,k} = 1$, $S_{k,k+1} = -1$ and 0 otherwise and $D = \text{diag}\{G_{1,1}, \ldots, G_{n,n}\}$: the limit distribution is $\chi_{n-1}$ when there is no cointegration and tends to 0 otherwise.
Having obtained evidence that the memory of the two processes is not different, the next step is a proper test of cointegration. Once again, both Robinson and Yajima (2002) and Marinucci and Robinson (2001) propose a semiparametric approach.

Marinucci and Robinson (2001) remark that since the matrix $G_{xx}$ is singular under cointegration, the joint estimation of the memory parameter following the minimization in (15) is not bound to give consistent estimates. They proposed to impose the condition

$$d_k = d_l = d_*$$

and to estimate the parameter using the loss function for the bivariate case: calling the estimate $\hat{d}_*$, they then considered the statistics

$$H_{km} = 8m(\hat{d}_* - \hat{d}_k)^2, \quad H_{lm} = 8m(\hat{d}_* - \hat{d}_l)^2,$$

arguing on the line of the Hausman test that these are asymptotically $\chi^2_1$ when there is no cointegration.

They also considered an interesting diagnostic for the case of cointegration, proposing to estimate the residuals with a consistent estimator and then discuss a conventional test on the order of integration.

The second test is from Robinson and Yajima: they remarked the fact that some of the eigenvalues of the matrix $G_{xx}$ should be zero under cointegration, so they proposed to estimate that matrix and compute the eigenvalues. Since $\hat{d}_*$ is not a consistent estimator in case of cointegration, they used

$$\hat{d}_* = (\hat{d}_1 + \ldots + \hat{d}_n)/n;$$

they also observed that if the same bandwidth is used for $\hat{d}_*$ and $\hat{G}_{xx}(\hat{d}_*, \ldots, \hat{d}_*)$ then these are perfectly correlated, so they suggested to compute $\hat{d}_*$ using another bandwidth $m_1$ that increases sufficiently fast to remove that effect. Letting $\hat{v}_1(\hat{d}_*), \ldots, \hat{v}_n(\hat{d}_*)$ be the ordered eigenvalues of $\hat{G}_{xx}(\hat{d}_*, \ldots, \hat{d}_*)$, Robinson and Yajima defined

$$s_j = \left( \hat{\sigma}_{n-j+1,n}^{(1)} \hat{\sigma}_{1,n}^{(1)} + \hat{\sigma}_{n-j+1,n}^{(2)} \hat{\sigma}_{1,n}^{(2)} \right)^{1/2}$$

and showed that when the rank of $G_{xx}$ is full

$$m^{1/2}(\hat{\pi}_j - \pi_j)/s_j \rightarrow_d N(0, 1) \text{ as } T \rightarrow \infty.$$

To test for the cointegration rank $r$ they considered

$$\hat{\pi}_r + z_a s_r/m^{1/2} \quad (19)$$
(where \( z_a \) is the critical value for the size \( a \)): evidence of cointegration is found if the computed value of the expression in (19) is below a pre-specified threshold (they suggested \( 0.1/n \)). They also proposed, as an alternative, to confront \( \hat{\pi}_r \) with another pre-specified threshold (they suggested \( 0.01/n \)).

Notice that Robinson and Yajima only formulated the test for variables having an order of integration below 0.5. Since we will be dealing with variables with a higher order of integration, we considered the alternative to take \( d_* \) fractional differences of the data, compute the periodogram of the series generated in this way and then estimate \( G_{xx} \) averaging the first \( m \) frequencies. This is exactly the procedure in Phillips and Ouliaris (1988), but allowing for fractional differencing: in practice, we replaced the periodogram in (11) with the one of the data differenced \( d_* \) times, and we set \( d_1 = 0, \ldots, d_n = 0 \).

If evidence of cointegration is obtained, Marinucci and Robinson (2001) estimated the cointegration parameter using a particular frequency domain least square (FDLS) statistic. The model they consider is very general, because no particular structure is imposed a part from some regularity conditions. They estimated the cointegrating parameter with

\[
\hat{\nu}_{m_2} = \hat{F}_{zy}(m_2)\hat{F}_{zy}(m_2),
\]

where

\[
\hat{F}_{zy}(m_2) = 2 \text{Re} \left\{ \frac{2\pi}{n} \sum_{j=1}^{m_2} I_{zy}(\lambda_j) \right\} - \frac{2\pi}{n} I_{zy}(\pi)1(m_2 = T/2)
\]

and

\[
\hat{F}_{zz}(m_2) = 2 \text{Re} \left\{ \frac{2\pi}{n} \sum_{j=1}^{m_2} I_{zz}(\lambda_j) \right\} - \frac{2\pi}{n} I_{zz}(\pi)1(m_2 = T/2).
\]

This corresponds to the OLS when \( m_2 = T/2 \), but it is in general different; Robinson and Marinucci (1999) also showed that, as long as the data are cointegrated and taking \( m_2 \) properly

\[
\frac{m_2}{T} + \frac{1}{m_2} \to 0 \quad \text{for} \quad T \to \infty,
\]

the FDLS is consistent even when the OLS is not, and that it is also more efficient than the OLS, in terms of a faster rate of consistency, in case the gap \( d - b \) is smaller than 1. They also suggested to compute the residuals of the cointegrating relation: despite the fact that these are not the true residuals, they “conjecture that asymptotic distributions” of the semiparametric estimates of the corresponding memory “are unaffected by the presence of estimated parameters, at least when FDLS are more-than-\( T^{1/2} \)-consistent”, a property that is in fact explicitly discussed by Hassler, Marmol and Velasco (2002) with respect to the log-periodogram regression estimate.

Since the expectations theory imposes a certain structure to the cointegration model, the semiparametric analysis, focussing on the parameters of interest only (the memory parameters, the cointegration rank, the cointegration parameter), is the natural approach. The semiparametric estimators are in fact faster to implement and, not requiring any specification of the short term components, they are not exposed to a potential
misspecification (there including a change in the short run dynamics, as hypothesized by Hansen) of a component which is not of interest.

In order to extend the analysis of the transmission of the policy impulses to the short run dynamics too, a fully parametric model must be specified. We then assume for the process in (13) a parametric linear structure

$$B(L; \varphi)u_t = \varepsilon_t$$

where

$$B(L; \varphi) = I_n + \sum_{j=1}^{\infty} B_j(\varphi)L^j$$

and $\varepsilon_t$ is an $n \times 1$ vector such that

$$E(\varepsilon_t \varepsilon_t') = \Omega E(\varepsilon_t \varepsilon_{t+s}') = 0$$

for any $s \neq 0$, with $\Omega$ positive definite, excluding any parametric restriction across the elements of $\Omega$ and $\nu, (d, b), \varphi$.

To test for cointegration using the procedure of Breitung and Hassler we introduce

$$v_t = \Delta^d x_t, \quad \widehat{\epsilon}_t = x_t - \widehat{A}_1 v_{t-1} - \ldots - \widehat{A}_p v_{t-p}$$

where $p$ is the order of a VAR for $\Delta^d z_t$ only, and

$$\widehat{\epsilon}_{t-1} = \sum_{j=1}^{t-p-1} j^{-1} \widehat{\epsilon}_{t-1-j}^*, \quad \widehat{\epsilon}_{t-1}^* = [\widehat{\epsilon}_{t-1}^*, v_{t-1}^*, \ldots, v_{t-p}^*]'$$

$$\tilde{\Sigma} = 1/T \sum_{t=1}^{T} \widehat{\epsilon}_{t-1}', \quad \tilde{S}_{10} = \sum_{t=1}^{T} \widetilde{\epsilon}_{t-1} \widehat{\epsilon}_{t-1}', \quad \tilde{S}_{11} = \sum_{t=1}^{T} \widetilde{\epsilon}_{t-1} \widehat{\epsilon}_{t-1}^*$$

we then have to compute the eigenvalues of

$$\lambda \tilde{\Sigma} - \tilde{S}_{10} (\tilde{S}_{11})^{-1} \tilde{S}_{10} = 0.$$

Breitung and Hassler tested the hypothesis of $r_0$ cointegrating vectors ordering the eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$ and computing the trace statistic $\sum_{j=1}^{n-r_0} \lambda_j$, which is asymptotically $\chi^2_{(n-r_0)}$. Since in practice $d$ is unknown, Breitung and Hassler suggested to replace it with a consistent estimate.

Upon knowing the cointegration rank, simultaneous estimation of both the short and the long run parameters can be quickly designed using a pseudo maximum likelihood approach.

When $\nu, d, b, \mu$ are known, we can compute $u_t$ and apply the filter $B(L; \theta)u_t$, which generates a vector time series $\varepsilon_t(\theta, \varphi)$: a gaussian type of loss function is
\[
\ln |\Omega| + \frac{1}{T} \sum_{j=1}^{T} \varepsilon_t(\theta, \varphi)' \Omega^{-1} \varepsilon_t(\theta, \varphi),
\]

and, concentrating \( \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(\theta, \varphi) \varepsilon_t(\theta, \varphi)' \) out, we can estimate \( \varphi \) minimizing
\[
\ln \left| \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(\theta, \varphi) \varepsilon_t(\theta, \varphi)' \right|.
\]

Similarly, if only \( \nu \) and \( \mu \) are known, it is possible to observe \( y_t - \nu' z_t - \mu y \) and then to compute \( \varepsilon(\theta_d, \theta_b, \theta, \varphi) = B(L; \theta, \varphi) u_t(\theta_d, \theta_e) \), with \( u_{y,t}(\theta_b) = \Delta^{\theta_b}(y_t - \nu' z_t - \mu y) \), \( u_{z,t}(\theta_d) = \Delta^{\theta_d}(z_t - \mu x) \). When \( \nu \) is unknown \( \varepsilon(\theta_d, \theta_b, \theta, \varphi) \) is computed for any possible \( \theta, \varphi \), and it is then indicated as \( \varepsilon(\theta, \theta_d, \theta_b, \theta, \varphi) \); finally, since \( \mu \) is unknown in practise we use mean corrected data, \( x_t - \bar{x} \) where \( \bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t \):
\[
u_t(\theta, \theta_d, \theta_b, \theta) = \begin{bmatrix}
\Delta^{\theta_b}(y_t - \bar{y} - \theta'_b (z_t - \bar{z})) \\
\Delta^{\theta_d}(z_t - \bar{z})
\end{bmatrix}.
\]

We then estimated \((\hat{\nu}, \hat{d}, \hat{b}, \hat{\varphi})\) as
\[
(\hat{\nu}, \hat{d}, \hat{b}, \hat{\varphi}) = \arg \min_{\theta, \theta_d, \theta_b, \theta, \varphi} L(\theta, \theta_d, \theta_b, \theta, \varphi)
\]
where
\[
L(\theta, \theta_d, \theta_b, \theta, \varphi) = \ln \left| \frac{1}{T - \kappa} \sum_{t=\kappa}^{T} \varepsilon_t(\theta, \theta_d, \theta_b, \theta, \varphi)' \varepsilon_t(\theta, \theta_d, \theta_b, \theta, \varphi) \right|.
\]

In practise, we also assume that only \( B_j, j \leq P < \infty \) are relevant, so the structure for \( u_t \) is a VAR(\( P \)), and setting \( \kappa > P \) it can be estimated by OLS; notice that in (45) we allow \( \kappa \) even larger than 1 because for low \( t \) the application of the filters \( \Delta^{\theta_b}, \Delta^{\theta_d} \) does not have many past observations (we set \( \kappa = P + 3 \)).

The maximization procedure in the bivariate example then involves only three groups of parameters \((\theta, \theta_d, \theta_b, \theta)\), and it follows these three steps:

1. “guess” a \( \theta, \theta_d, \theta_b, \theta \),
2. compute the residual
\[
\varepsilon_t(\nu, \hat{\mu}_y) = y_t - \bar{y} - \theta'_b (z_t - \bar{z})
\]
for the candidate \( \theta, \nu \);
3. for each \( \theta, \nu \), consider
\[
u_{y,t}(\theta, \theta_b, \hat{\mu}_y) = \Delta^{\theta_b} \varepsilon_t(\theta, \nu, \hat{\mu}_y)
\]
for the candidate \( \theta_b \) and
\[
u_{z,t}(\theta_d, \hat{\mu}_x) = \Delta^{\theta_d}(z_t - \bar{z})
\]
for the candidate \( \theta_d; \varepsilon_t(\theta, \theta_d, \theta_b, \theta, \varphi, \hat{\mu}) \) are then computed as the OLS residuals of a VAR(\( P \)) formulated for \( u_t(\theta, \theta_d, \theta_b, \theta) \).
If the model is correctly specified, and given other regularity conditions as in Appendix B, it also follows that

\[(T - \kappa)(L(\tilde{\nu}, \hat{d}, \hat{b}, \tilde{\varphi}) - L(\nu, d, b, \varphi)) \to_d \chi^2(K) \text{ as } T \to \infty\]

where \(K\) is the number of elements of the vector \((vec(\nu)', d, b, \varphi)')\; linear restrictions on a subset of \((vec(\nu)', d, b, \varphi)')\ are dealt with in a similar way, and as usual the degrees of freedom in the limit distribution correspond to the number of restrictions tested.

As already observed in Johansen (1991) (and by making use of exactly the same argument), if \(d - b > 1/2\) one can conduct inference on \(\tilde{\nu}\) tacking \(d, b, \varphi\) as known and viceversa; the same applies to the \(\chi^2\) test: the test statistic decomposes into a test for \(vec(\tilde{\nu})\) and an independent test for \((\hat{d}, \hat{b}, \tilde{\varphi}')'\).

4 The Term Structure of Interest Rates, an Empirical Analysis

We applied the fractional model to the 1, 3, 6 and 12 months interest rates on the London interbank deposit in US$ (monthly averages of the offer rate) over the period 10/1979 to 01/2002 (included), corresponding to the DataStream identification codes USI60LDC, USI60LDD, USI60LDE, USI60LDF. The period was selected because Clarida Gali and Gertler (2000) suggested that with the appointment of Volker as a chairman the Fed took a more aggressive attitude towards inflation, and indeed a break in that point was suspected in that point in many applied analyses. The London InterBank Offer Rate is the typical measure of the cost of funds in the US, because it is not affected by any regulation imposed by the central bank; in fact for the same argument the LIBOR had been in the past a good measure of the effective cost of funds for several European countries too. Since these are monthly data, using the notation above they are \(i_{t,t+1}, i_{t,t+3}, i_{t,t+6}\) and \(i_{t,t+12}\), but we shortened it to \(i_{1t}, i_{3t}, i_{6t}, i_{12t}\).

A key element of the semiparametric analysis is the choice of the bandwidth \(m\): the earliest semiparametric applications used \(T^{0.5}\) or \(T^{0.6}\), while Henry and Robinson (1996) showed that the local Whittle estimator that the MSE is minimized setting \(m\) proportional to \(T^{4/5}\). In the Appendix A we followed the same approach, deriving the formulas to compute the bandwidths that are optimal in the minimum MSE sense for the local Whittle estimator when the Fourier transforms are tapered.

In our semiparametric analysis anyway we decided not to rely on minimal MSE only, because the approximation of the (pseudo) spectral density with \(G\lambda^{-2d}\) only holds in a reasonably small neighborhood of \(\lambda = 0\) and results based on asymptotic theory are then less reliable the larger \(m\) is. Since in our application the semiparametric analysis was only introduced to obtain preliminary evidence, it may be preferable to use smaller bandwidths as well, then, especially if the tests lead to convincing conclusions anyway. Another possible disadvantage of the automatic selection of the optimal bandwidth is that a certain sensitivity to the starting value in the iterative procedure can be found,
when the sample size is relatively small. We then coupled the bandwidths selected with minimum MSE procedures with the user chosen $m = 25$ (when the raw periodograms were used) and $m = 30$ (when tapering was applied), and presented the results for both. These user chosen bandwidths are comparable in the sense that they are both optimal for an AR(1) structure with autoregressive coefficient of 0.4 for the short memory component, albeit of course we did not assume such a structure and only chose them because in simulations with samples of a similar dimension the dispersion of the estimates did not exceed too much the one prescribed in the asymptotic theory.

The plot of $i1$ and $i12$ is in Figure 1.

A naive inspection of the plot suggested that the data are indeed reverting to a potential mean only very slowly if at all, and the apparent downward trend in the first part of the period can be interpreted as the movement towards the long run equilibrium after a particularly large perturbation; in alternative, we did not rule out the possibility that $d = 1$, as in the mainstream cointegration literature. We then had to choose whether to estimate the memory parameter using the original data or the first differences, and for this purpose a preliminary estimation using tapering was run. The bandwidths chosen with the automatic procedure were rather large, to the point that in two cases they selected all the frequencies up to $T/2$, our upper limit to $m$, and we think that the results from the more conservative $m = 30$ are more reliable. In any case, there are not relevant differences: the estimates, presented in Table 1, are indeed in the set $(0.5, 1.5)$, justifying the analysis of the first difference of the data.

We then estimated $d$ using the standard local Whittle procedure on the first difference of the data, collecting the results in Table 2 (notice that the estimates computed from the first differences are indicated as $d_{\Delta i}$, while those computed from the original, raw data are indicated as $\tilde{d}_i$). The bandwidths selected with the automatic procedure were lower than the corresponding ones for tapered data, and the estimates for $d_{\Delta i}$ were between $-0.32$ and $-0.16$. This wide dispersion partially depended on a certain instability of the estimated values with respect to the change of bandwidth: setting $m = 25$ the range was reduced to $-0.14$ to $-0.05$. We also considered a bandwidth optimal for the vector of the four rates, as discussed in Appendix A: we obtained $m = 48$, and an even smaller range, covering $-0.24$ to $-0.18$ only. Point estimates of $d$ were then below 1, albeit not significantly so for the conservative bandwidth $m = 25$.

To justify the $I(0)$ structure despite the extensive evidence in the literature against it based on the Dickey and Fuller test, it was often argued that the power of the test is very low. Having estimated the order of integration directly, we treated both the $I(0)$ and $I(1)$ specifications in the same way, and the evidence against the $I(0)$ model is then far more convincing. Considering our estimated order of integration, it is not surprising that the conventional unit root tests were in favour of the $I(1)$ model: even assuming that the data were neither $I(0)$ nor $I(1)$, they appeared to be far closer to being $I(1)$, and the Dickey and Fuller test just reflected this fact.

We next tested the hypothesis that the memory is the same for all the series, presenting in Table 3 the squares of the pairwise statistics $T$. As before, we selected a bandwidth that was optimal for the two estimates, and added the case $m = 25$: we rejected the null hypothesis of a common order of integration in two combinations out
of six with automatic bandwidth and never with \( m = 25 \). It is fair to suspect that the two rejections observed were due to the fact that the distribution is not well defined if the data are cointegrated, but, rather than specifying the additional term \( h(T) \) and considering \( \tilde{T} \), we tested the hypothesis jointly using the formula (18) and a \( \chi^2_3 \) limit distribution. Since we set \( h(T) = 0 \), this is analogue to the formula based on \( \tilde{T} \), so again a failure to reject the null implies a similar result for any \( h(T) \) in (18): the computed test statistics were 2.93 for the optimal bandwidth \( (m = 48) \) and 2.74 for \( m = 25 \), providing evidence in favour of a common order of integration. To estimate the memory of the data more efficiently we then pooled the four individual estimates: we obtained \( \hat{d}_{*,\Delta i} + 1 = 0.83 \) using \( m = 48 \) and \( \hat{d}_{*,\Delta i} + 1 = 0.91 \) with \( m = 25 \).

Having obtained evidence in favour of a common order of integration, we tested for cointegration. The obvious candidate cointegration rank is \( n - 1 \), since this is the one required under the expectations theory and it is also used in many empirical analyses. The previous finding that the order of integration is the same for all the data is indeed consistent with this assumption: we then proceed to test if pairs of interest rates with the Robinson and Marinucci approach. If \( d \) is jointly estimated, an optimal bandwidth for a multivariate local Whittle estimator can be derived using the same approach of Henry and Robinson (see Appendix A for details). The results of the tests are presented in Table 4: we rejected the null of no cointegration in 5 combinations out of 12; we also had 9 rejections (out of 12) when \( m = 25 \) was applied. In all the 24 combinations we always had that \( \hat{d}_* \) is lower than both the orders individually estimated for the two series, which makes us suspect that the failure to reject in the remaining cases may be due to a type II error. The results of the tests proposed in Robinson and Yajima, presented in Table 5a, are similar, maybe leaving a little more room for the interpretation. The optimal bandwidth in this case was selected with the goal of minimizing the MSE of the vector of the memory parameters of the four rates: it turned out to be \( m = 48 \), and we set \( m_1 = 52 \) to estimate \( G_{xx} \). Although the test can be designed for any cointegration rank, we discussed \( r = 3 \) only, because this is the one relevant for the theory. The maximum of the rescaled sum of the \( j \) eigenvalues, \( \hat{\pi}_3 \), was about 0.02 for the exactly differenced data, larger than the suggested threshold 0.01/\( n = 0.0025 \); the statistic \( \hat{\pi}_3 + z_a s_3/m^{1/2} \) on the other hand was about at the level of the threshold 0.1/\( n = 0.025 \). Slightly better results followed for \( m = 25 \) and \( m_1 = 27 \): \( \hat{\pi}_3 \) was 0.009, still above the threshold but closer to it, while \( \hat{\pi}_3 + z_a s_3/m^{1/2} \) took value 0.012, well below the corresponding threshold. We also considered testing only couple of rates, as in the Marinucci and Robinson approach: the theory suggests \( r = 1 \) in that case, so in Table 5b we presented the statistics corresponding to these tests only. The evidence was still mixed for the \( \hat{\pi}_1 \) statistics, pointing at cointegration when the test is run for rates that are contiguous \((i_1, \ i_3, \ i_3, \ and \ i_6t, \ i_6t, \ and \ i_2t)\) but not otherwise, resulting in 6 rejections out 12 using the threshold 0.01/\( n = 0.005 \); once again, the rejections were less (only 4 out 12) for the statistics \( \hat{\pi}_1 + z_a s_1/m^{1/2} \).

The model we estimated is then
\[
\begin{align*}
    i_{1t} &= \mu_1 + \Delta^{-d}u_{1t}1(t > 0) \\
    i_{3t} &= \mu_3 + \nu_3 i_{1t} + \Delta^{-b_3}u_{3t}1(t > 0) \\
    i_{6t} &= \mu_6 + \nu_6 i_{1t} + \Delta^{-b_6}u_{6t}1(t > 0) \\
    i_{12t} &= \mu_{12} + \nu_{12} i_{1t} + \Delta^{-b_{12}}u_{12t}1(t > 0).
\end{align*}
\]

Since the shorter rate is the closest to the Federal Funds rate, these combinations are informative about the existence of a long run relation for the transmission of the monetary impulses.

For the Narrow Band FDLS we observed from the simulations in Marinucci and Robinson that even a moderate correlation in the innovations \(\eta_{1,t}, \eta_{2,t}\) can induce a relevant bias in the OLS estimate in small samples, so that the MSE is lower when only a little number of frequencies is considered. Since in our data approximately 72% of the total variation was concentrated in the frequencies up to \(\frac{2\pi^3}{T}\), and up to 80% by the frequency \(\frac{2\pi^5}{T}\), we set \(m_2 = 5\) for the computation of \(\hat{\nu}_{m_2}\). We reported the results in Table 6: we found that the correction of the Narrow Band FDLS on the OLS was very little, as if either the gap \(d - b\) was very wide or the correlation between the innovations and the explanatory variables was not large. The estimates were slightly decreasing with the increase of the difference in the maturities, the minimum of 0.945 corresponding to the relation between \(i_{1t}\) and \(i_{12t}\).

The memory of the residuals ranged between 0.18 and 0.34 when estimated with the optimal bandwidth, and between 0.22 and 0.44 using \(m = 25\). This outcome mirrored the one of the pairwise cointegration tests, because the results in favour of cointegration between \(i_{1t}\) and \(i_{3t}\) and between \(i_{1t}\) and \(i_{6t}\) were tighter than the one for \(i_{1t}\) and \(i_{12t}\).

We conclude the semiparametric analysis discussing the implication of these results on the expectations theory and on the validity of the interest rate spread as a long run relation: the cointegrating parameters should be 1, and the semiparametric estimates were indeed reasonably close to it, at least in the first two cases. To further support this remark, we also notice that the estimate of the memory parameters \((i_{12} - i_{1}), (i_{6} - i_{1}), (i_{3} - i_{1})\) and of \((i_{12} - \hat{\nu}_{m_2}i_{1}), (i_{6} - \hat{\nu}_{m_2}i_{1}), (i_{3} - \hat{\nu}_{m_2}i_{1})\) are basically the same ones. The expectations theory anyway requires not only \(\nu = 1\) but also that \(b = 0\) for all the combinations, so we analysed this joint hypothesis testing \(H_0 : \{b = 0\}\) against \(H_0 : \{b > 0\}\) for the pairs \((i_{12} - i_{1}), (i_{6} - i_{1}), (i_{3} - i_{1})\), and we found that the order of the residuals was always too large, causing the rejection of the null hypothesis in all the cases. Evidence of failure of the expectations theory is common in the literature, but it was mainly observed though particular reparameterizations of the short run dynamics: an \(I(0)\) structure for the residuals was usually been taken for granted and not questioned before. Our results on the other hand showed that when fractional integration was allowed for, then evidence of residual strong memory appeared.

Despite the rejection of the expectations theory, the parametric analysis is still interesting because it is motivated by the study of the transmission mechanism of monetary impulses. The semiparametric analysis only gave answers for the reaction in the long
run, but the central bank and the financial markets operate considering the timing with great care. Of particular importance is also to assess whether the long term relations can indeed be represented as interest rate spreads, or if the effect of monetary policy impulses damps along the term structure.

Furthermore, even if the medium term rates are not anticipating the future short ones exactly, they could have some information: despite the rejection of the expectations theory, Campbell and Shiller (1987) think that “the spread predicts short run movements almost correctly”, and a sudden increase in the long rates, for example, may prelude to a tightening of the monetary policy, even though we can not rely on the expectations theory to quantify the exact measure of the future intervention.

For the cointegration test and for the estimation of the parametric model, we first had to describe the short term dynamics specifying the matrices $B(L, \varphi)$: since we assumed a VAR($P$), we only had to determine the number of the lags. For this purpose we computed the residuals $u_{1,t}(\Delta y_{1,t} + 1, \mu_1)$, $u_{3,t}(\Delta y_{3,t}, \mu_3)$, $u_{6,t}(\Delta y_{6,t}, \mu_6)$, $u_{12,t}(\Delta y_{12,t}, \mu_{12})$ using the semiparametric consistent estimates, and we applied the Schwartz (SC) and Hannan and Quinn (HQ) information criteria to these series, discarding the first two observations as in (45). In all the cases a VAR(2) was selected, and, in order to drop the first observations, we set $\kappa = 5$ in the estimations.

Notice that in the cointegration test preliminary knowledge of $d$ is also required: Hassler and Breitung remark that it can be replaced with a consistent estimate, but in line with the spirit of the parametric model, we fitted an ARFIMA(2,$d$,0) to the first differences of each rate, averaged the estimates of each memory parameter and added back 1, obtaining 0.86. The cointegration test replicated the conflicting outcome of the semiparametric test: once again we found rank 2 using the whole vector and the 5% test (but the realization of the test statistic was very close to the critical value, and rank 3 would have followed taking size 10%), while testing only couple of rates we always found evidence of cointegration, thus pointing at rank 3 in the whole vector. We collected the results in Table 8a and 8b.

Having set the cointegration rank to 3, we finally proceeded to the estimation of the parametric model. The estimates of the long run parameters were

<table>
<thead>
<tr>
<th>$\hat{\nu}_3$</th>
<th>$\hat{\nu}_6$</th>
<th>$\hat{\nu}_{12}$</th>
<th>$\hat{b}_3$</th>
<th>$\hat{b}_6$</th>
<th>$\hat{b}_{12}$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>1.01</td>
<td>0.98</td>
<td>0.19</td>
<td>0.21</td>
<td>0.25</td>
<td>0.86</td>
</tr>
</tbody>
</table>

as we also reported in Table 9; the estimates of the short run parameters for $B(L, \varphi)$ (as in 41) are in Table 10 and those for $\Omega$ in Table 11.

The estimates of the long run parameters resembled those from the semiparametric analysis: $\hat{\nu}_{12}$ and $\hat{\nu}_6$ were slightly closer to the parameters prescribed from the economic theory, and a larger correction took place for $\hat{b}_{12}$. The estimated cointegrating parameters were very close to 1, indeed even closer than the narrow band FDLS, and the effect of cointegration was quite relevant, the gaps $\hat{d} - \hat{b}$ ranging between 0.61 and 0.67. Since the asymptotic distributions of the estimates $\hat{\nu}$ and the likelihood ratio tests involving both the cointegration parameters and the other ones depend on the size of that gap, we first tested

$$H_0 : \{d - b_3 = 0.49\} \ v.s. \ H_1 : \{d - b_3 \neq 0.49\}$$
Each likelihood ratio test has a $\chi^2$ distribution (under the null), and the computed test statistics were 0.93, 0.68 and 0.38 respectively: despite the fact that the point estimates yielded gaps $d - b$ larger than 1/2, the difference was not statistically significant, so a simultaneous fully parametric approach may have some advantage on the two step ML approach described by Robinson and Hualde with these data.

We next tested the structural hypothesis that all the cointegrating parameters were 1, as it is usually assumed in the economic theory:

$$H_0: \{\nu_3 = 1, \nu_6 = 1, \nu_{12} = 1\} \text{ v.s. } H_1: \{\nu_3 \neq 1 \&/\ or \ \nu_6 = 1 \&/\ or \ \nu_{12} = 1\},$$

The corresponding test statistic took value 7.5, just below the critical value 7.8 (5% test for $\chi^2_3$ distribution), so the hypothesis was not rejected. The two other relevant hypothesis concerned the order of integration of the data: the expectation theory required

$$H_0: \{b_3 = 0, b_6 = 0, b_{12} = 0\} \text{ v.s. } H_1: \{b_3 \neq 0 \&/\ or \ b_6 = 0 \&/\ or \ b_{12} = 0\},$$

but the hypothesis was rejected (the test statistic took value 8.3); on the other hand, we again failed to provide a convincing evidence against the hypothesis of an unit root:

$$H_0: \{d = 1\} \text{ v.s. } H_1: \{d \neq 1\}$$

had a realized test statistic of 1.7.

As in the semiparametric analysis, the estimates of the orders of the residuals maintained the property that the closer the maturities are, the faster the adjustment, but this feature was not statistically significant: testing

$$H_0: \{b_3 = b_6 = b_{12}\} \text{ v.s. } H_1: \{b_3 \neq b_6 \&/\ or \ b_6 \neq b_{12}\},$$

the computed statistic took value 1.3, far below the critical value 6.0: the estimate of the memory of the residuals under this restriction was 0.2.

For a deeper structural analysis it is interesting to provide an interpretation of the parameters in the VAR model. Since the estimated cointegrating coefficients were very close to 1, these equations were similar to the spread between the long and the short rate, and in this sense we interpreted them in the dynamic analysis.

A simulation of the dynamic effects of an innovation in $i_1$ is clearly important to analyse the first step of the transmission mechanism of monetary policy. Despite the rejection of the expectations theory, it is also interesting to observe whether the spreads and the long term rates are useful to predict the short term one as Campbell and Shiller
suggested. To address these issues we computed the structuralised impulse response functions of the estimated model.

We then built the structuralised impulse response function assuming that the contemporaneous correlation moved from the shortest to the longest maturity, as we argued it happens for the unexpected impulses of monetary policy. The sample covariances of the residuals seemed to confirm this assumption: in fact, the residuals of the cointegrating relations were all negatively correlated to the ones of the interest rates process, and were positively correlated between themselves. This means that the spreads actually decreased when the 1 month rate increased: the longer rates followed the innovation in the short rate, but the contemporaneous variation was smaller, and it also got smaller when the maturity of the contract was longer. The positive correlation of the estimated innovations of the spreads suggested that all the rates reacted to some pieces of news occurred during the month about the future short term rates, anticipating the change; the correlation was also very strong, as if the whole term structure reacted to the same type of information.

All the estimated cointegrating residuals appeared to anticipate a future increase of the interest rate, showing that the contracts with longer maturities successfully embodied a piece of information about the future, as we see from Figure 2.

We also plotted in Figure 3 the estimated reaction of the three rates with respect to an innovation to the short term rate: the estimated effect of the innovation was temporary, as we can see form the fact that after 24 periods it was still decreasing, but it is also interesting to observe that the one month rate was still estimated to be 69 basis points above the starting value. The estimated reaction of the longer rates to a 100 basis points increase in the short rate were 93, 86 and 72 respectively, while the average difference of the short rate with respect to the original levels were 114, 102 and 91 basis points for 3, 6, and 12 periods after the innovation (including the one in which the innovation took place): the long rates underreacted slightly but they appear to be informative with respect to the future short ones.

5 Conclusion

Fractional integration and cointegration allow for a more flexible description of the characteristics of an economic time series. We consider here an example in which by restricting the attention to $I(1)$ and $I(0)$ models only we would implicitly exclude some properties that are required by the economic theory. Fractional models allow to reconcile the apparently alternative theories. We present a multivariate model for 4 US interest rates for different maturities to study the implications of fractional integration and cointegration on the expectations theory for interest rates, on the transmission mechanism of monetary policy, and on the stabilization of inflation more in general.

The first, semiparametric, analysis was sufficient to rule out the expectation theory, but it provided evidence in favour of the existence of a long run relation as required for the transmission of monetary policy, possibly taking the form of the interest rate spread. The fully parametric analysis confirmed this conclusion, and indicated that the spreads are informative with respect to the future rates. Evidence of long memory in the data,
obtained in the semiparametric analysis, was confirmed by the parametric specification: point estimates indicated a fractional dynamics for the data, although we were not able to reject the hypothesis of an $I(1)$ process.

**Appendix A. Selection of Optimal Bandwidth**

*Optimal bandwidth: the standard case:*

Henry and Robinson extended the assumption on the spectral density to

$$f(\lambda) \sim G\lambda^{-d} (1 + E_b(d)\lambda^b + o(\lambda^b)) \text{ as } \lambda \to 0^+,$$

(A.1)

and obtained for the MSE the formula

$$\frac{1}{4m} \left[ \frac{1}{m} + \frac{b^2 E_b(d)^2}{(b + 1)^4} \lambda^{2b} \right],$$

(A.2)

and the optimal bandwidth

$$m_{opt} = \left[ \frac{(b + 1)^4}{2b^2 E_b(d)^2 (2\pi)^{2b}} \right]^{\frac{1}{1+2b}} T^{\frac{2b}{1+2b}}.
$$

(A.3)

Setting the smoothest specification ($b = 2$, an assumption that is quite common in the literature because it corresponds to the case of an ARMA structure for the short memory component) the optimal bandwidth is $m_{opt} = (\frac{3T}{4\pi})^{1/5} |E_2(d)|^{-2/5}$. The term $E_b(d)$ can be estimated, given a bandwidth $m$ and knowledge of $G$ and $d$, as the coefficient of $\lambda^2$ in the regression of $I_{\{(G\lambda)^2\}}$ on 1 and $\lambda^2$ for $j = 1$ to $m$. Since anyway $m_{opt}$, $G$ and $d$ are unknown, Henry and Robinson proposed an iterative procedure, in which an initial $m^{(0)} \propto n^{4/5}$ is set (preliminary simulations suggested us to set $m^{(0)} = 1/2n^{4/5}$), allowing a first step estimation of $G$ and $d$ and then of $E_2(d_0)$. The estimates of $d$, $G$ and $E_2(d)$ are then updated replacing $m^{(0)}$ with $\hat{m}^{(1)}$, and iterating the procedure until convergence.

*Optimal bandwidth: tapered periodogram:*

When tapering is considered, we replace the MSE formula with

$$\frac{1}{4m} \left[ \frac{\Phi P}{m} + \frac{b^2}{(b + 1)^4} \lambda^{2b} \right],$$

(A.4)

and the optimal bandwidth with

$$m_{opt} = \left[ \frac{\Phi P (b + 1)^4}{2b^2 E_b(d)^2 (2\pi)^{2b}} \right]^{\frac{1}{1+2b}} T^{\frac{2b}{1+2b}}.
$$

(A.5)

*Optimal bandwidth: vector of separate estimates:*

When a vector
\[ d = (d_1, d_2)' \]

is considered, an optimal bandwidth is obtained from

\[
\min_m E \| \tilde{d} - d_0 \|^2 ,
\]

\[
\min_m (\text{var}(d_2) + \text{var}(d_2) + \text{bias}^2(d_2) + \text{bias}^2(d_2)),
\]

where we can use \(m \text{var}(d_2) \to 1/4, m \text{var}(d_2) \to 1/4\).

Applying the arguments of Henry and Robinson, the MSE formula then becomes

\[
\frac{2}{4} \left[ \frac{1}{m} + \frac{b^2}{(b+1)^4} \lambda_m^{2b E_b(d_1)^2 + E_b(d_2)^2} \right]
\]

and the optimal bandwidth

\[
\hat{E}(d) = \frac{E_b(d_1)^2 + E_b(d_2)^2}{2}, m_{opt} = \left[ \frac{(b + 1)^4}{2b^3 E(d)(2\pi)^{2b}} \right]^{1/2b} T^{1/2b}.
\]

**Optimal bandwidth: joint simultaneous estimates:**

The minimal MSE criterion can also be used to obtain an optimal bandwidth when dealing with a multivariate process. This is of interest in the application of the cointegration test described in Marinucci and Robinson 2001, where the memory parameter is estimated minimizing

\[
Q(d_*) = -2(2d_*) \frac{1}{m} \sum_j \ln \lambda_j + \ln \left| \hat{G}(d_*) \right|
\]

and

\[
\hat{G}(d_*) = \frac{1}{m} \sum_j \lambda_j^{2d_*} I(\lambda_j),
\]

which is the loss function considered by Lobato (1999) with the additional constraint that \(d_1 = d_2 = d_*\).

To generalize the spectral density we assumed

\[
f(\lambda) = \lambda^{-2d} (G + E_b(d)\lambda^b + o(\lambda^2)) \quad \text{as } \lambda \to 0^+,
\]

where

\[
E_b(d) = A_b(d)G A_b(d)'
\]

and, as in Lobato’s work, we assumed that \(G\) is symmetric, real and positive definite.

To obtain the MSE in this case, we first computed

\[
\frac{\partial Q(d_*)}{\partial d_*} = \frac{2}{m} \sum_j \nu_j T r(\hat{G}^{-1} \lambda_j^{2d_*} I(\lambda_j))
\]
where

\[ \nu_j = \ln \lambda_j - \frac{1}{m} \sum_{j} \ln \lambda_j . \]  

(A.14)

Replacing \( I(\lambda_j) \) with \( f(\lambda_j) \) and \( \hat{G} \) in \( G \) in A.20, and computing it in \( d \),

\[ \frac{\partial Q(d_*)}{\partial d_*}|_{d_*=d_0} \sim_p \frac{2}{m} \sum_{j} \nu_j \text{Tr}(G^{-1} E_b(d) \lambda_j^b) = 2C(d) \frac{b}{(1+b)^2} \lambda_m^b, \]  

(A.15)

where we set

\[ C(d) = \text{Tr}(G^{-1} E_b(d_*)). \]  

(A.16)

As already noticed by Marinucci and Robinson, the asymptotic variance of \( \bar{d}_* \) is

\[ m \text{Var}(\bar{d}_*) \to 1/8, \]  

(A.17)

and \( \left( \frac{\partial^2 Q(d_*)}{(\partial d_*)^2} \right)|_{d_*=d_0} \to 1/8 \), so the bias is

\[ \text{bias} = \frac{1}{8} 2C(d) \frac{b}{(1+b)^2} \lambda_m^b = \frac{1}{4} C(d) \frac{b}{(1+b)^2} \lambda_m^b, \]  

(A.18)

the MSE is

\[ MSE = \left[ \frac{1}{8m} + \left( \frac{1}{4} C(d) \frac{b}{(1+b)^2} \lambda_m^b \right)^2 \right] \]  

(A.19)

and

\[ m_{opt} = \left[ \frac{1}{C(d)^2} \left( \frac{1+b}{b^2} \right)^4 \left( \frac{3T}{2\pi} \right)^4 \frac{1}{8} \right]^{\frac{1}{8}}, \]  

(A.20)

which is

\[ m_{opt} = \left[ \frac{1}{C(d)^2} \left( \frac{3T}{2\pi} \right)^4 \frac{1}{8} \right]^{\frac{1}{8}} \]  

(A.21)

when \( b = 2 \) is replaced.

As in the univariate case discussed by Henry and Robinson, this bandwidth is actually infeasible, because \( C(d) \) is unknown. I conjecture one can estimate it running the regressions of \( \lambda_j^2 \) on \( 1 \) and \( \lambda_j^2 \), obtaining the elements in the position \( k,l \) of \( \hat{G} \) and of \( \hat{E}_b(d_*) \), and computing \( \hat{C}(d_*) = \text{Tr}((\hat{G})^{-1} \hat{E}_b(d_*)) \).

**Optimal bandwidth for the Marinucci and Robinson cointegration test:**

We can also compute the optimal bandwidth for the test for cointegration discussed by Marinucci and Robinson: they remark that when the time series are not cointegrated, then

\[ 8m(\bar{d}_k - \bar{d}_*)^2 \rightarrow_d \chi(1), \]  

24
arguing that the null hypothesis of no cointegration can be rejected if the computed test statistic exceeds the critical value. An optimal bandwidth can then be computed applying the MSE criterion to the vector \((\delta_k, \overline{d_*})'\): in this case it is

\[
MSE = \left[ \frac{1}{4m} + \frac{1}{4} \left( \frac{b^2}{(b+1)^4} \right) E_b(d_k)^2 \lambda_m^{2b} + \frac{1}{8m} + \frac{1}{16} C(d_*)^2 \frac{b^2}{(1+b)^4} \lambda_m^{2b} \right],
\]

so

\[
m_{opt} = \left[ \frac{3}{2} \left( \frac{1}{4 \bar{C}} \right) b^2 \left( \frac{T}{2\pi} \right)^4 \right]^{\frac{1}{4T}} \text{ where } \bar{C} = E_b(d_k)^2 + \frac{1}{4} C(d_*)^2 \tag{A.22}
\]

\[
m_{opt} = \left[ \frac{3}{2} \left( \frac{1}{4 \bar{C}} \right) \left( \frac{3T}{4\pi} \right)^4 \right]^{\frac{1}{2}} \tag{A.23}
\]

when \(b = 2\).

**Appendix B. Sufficient Conditions for Maximum Likelihood Estimation**

Considering the model (12), the structure in (20) is formalized as

\[
E(u_t) = 0, \quad E(u_t u_{t+j}') = \int_{-\pi}^{\pi} f_{uu}(e^{i\lambda}) e^{i\lambda j} d\lambda \tag{B.1}
\]

and \(f_{uu}(e^{i\lambda})\) is continuous and nonzero at all frequencies.

Robinson and Hualde assumed a moving average structure as in

\[
u_t = A(L)\varepsilon_t, \tag{B.2}
\]

where

\[
A(s) = I_n + \sum_{j=1}^{\infty} A_j s^j, \tag{B.3}
\]

and \(\varepsilon_t\) as in (21).

In general the matrices in \(A(s)\) and in \(\Omega\) are known functions of the unknown vector of parameters \(\varphi, \omega\) and are indicated as \(A(s; \varphi)\) and \(\Omega(\omega)\). We specify our assumption with respect to the generic matrices \(A(s; \varphi)\) and \(\Omega(\omega)\), and to \(f_{uu}(e^{i\lambda}; \varphi, \omega)\), defined as

\[
f_{uu}(e^{i\lambda}; \varphi, \omega) = \frac{1}{2\pi} A(e^{i\lambda}; \varphi) \Omega(\omega) A(e^{-i\lambda}; \varphi)'.
\]

We also let

\[
\theta = (\text{vec}(\theta_\omega), \theta_d, \theta_b, \theta_\varphi)', \quad \theta_0 = (\text{vec}(\nu)', d, b, \varphi)'.
\]
Assumption 1
(i) $\Theta$ is a compact subset of $R^{\dim(\theta)}$, $\theta \in \Theta$ and $\theta_0 \in \Theta$; $S$ is a compact subset of $R^{\dim(\omega)}$, $\omega_0 \in S$ and $\omega \in S$;
(ii) $A(e^{i\lambda}; \theta_\varphi)$ is continuous at all $(\lambda, \theta_\varphi)$, and $\det(A(s)) \neq 0$ for $|s| = 1$;
(iii) $\Omega(\theta_\omega)$ is continuous for all $\theta_\omega$ and positive definite;
(iv) $\theta_\varphi \neq \theta_\psi$, implies $A(e^{i\lambda}; \theta_\varphi) \neq A(e^{i\lambda}; \theta_\psi)$ on a set of positive measure; $\theta_{\omega_1} \neq \theta_{\omega_2}$ implies $\Omega(\theta_{\omega_1}) \neq \Omega(\theta_{\omega_2})$ on a set of positive measure.
(v) $\sum_{j=0}^{\infty} \|A_j\|^2 < \infty$.

Assumption 2
(i) $f_{uu}(e^{i\lambda}; \theta_\varphi, \theta_\omega)$ is differentiable in $\lambda$, with derivatives bounded and continuous at all $(\lambda; \theta_\varphi, \theta_\omega)$.
(ii) $f_{uu}(e^{i\lambda}; \theta_\varphi, \theta_\omega)$ is differentiable in $(\theta_\varphi, \theta_\omega)$, with derivatives bounded and continuous at all $(\lambda; \theta_\varphi, \theta_\omega)$ and differentiable in $\lambda$ at all $(\lambda; \theta_\varphi, \theta_\omega)$.

Assumption 3
$A(\lambda)$ is differentiable in $\lambda$ with derivative in $\text{Lip}(\eta)$ with $\eta > 1/2$;

Assumption 4
(i) $E|\varepsilon_i|^q < \infty$: for $q = 4$ if $d - b < 1/2$, and for some $q > 4$, $q > 2/(2|d - b| - 1)$ if $d - b > 1/2$.
(ii) $\sum_{j=0}^{\infty} j \|A_j\|^2 < \infty$.

Assumption 5
(i) $f_{uu}(e^{i\lambda}; \theta_\varphi, \theta_\omega)$ has two bounded and continuous derivatives with respect to $\theta_\varphi, \theta_\omega$ at all $(\lambda; \theta_\varphi, \theta_\omega)$, and these derivatives are differentiable with bounded and continuous derivative with respect to $\lambda$ at all $(\lambda; \theta_\varphi, \theta_\omega)$;
(ii) $\theta_\omega$ is parametrized separately from $\theta$.

Letting

$$M(\nu) = \begin{bmatrix} I_r & -\nu' \\ 0 & I_{n-r} \end{bmatrix}, \quad D(z; d, b) = \begin{bmatrix} (1 - z)^b I_r & 0 \\ 0 & (1 - z)^d I_{n-r} \end{bmatrix},$$

(B.4)

(where again the additional assumption $b_1 = \ldots = b_r = b$ is only introduced to shorten the notation) and

$$\Upsilon(z; \theta) = M^{-1}(\theta_\nu)D(z; \theta_d, \theta_b)A(z; \theta_\varphi),$$

(B.5)
consider the model in (12), (B.1) - (B.3): under the Assumptions 1 to 5 and
\[ d - b < 1/2 \]  
then \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \), and
\[ n^{1/2}(\hat{\theta} - \theta_0) \to_d N(0, \Sigma) \]
where the element in position \( l, k \) for \( l, k = 1..\dim(\theta) \) is
\[ (\Sigma)_{l,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr} \left\{ (Y(e^{i\lambda})^{-1} \frac{\partial Y(e^{i\lambda}; \theta)}{\partial \theta_l} \Omega (Y(e^{-i\lambda})^{-1} \frac{\partial Y(e^{-i\lambda}; \theta)}{\partial \theta_k})' \Omega^{-1} \big|_{\theta=\theta_0} \right\} d\lambda. \]

When the “cointegration gap” \( d - b \) exceeds \( 1/2 \), the asymptotic distribution follows from the one in Robinson and Hualde. Introduce \( W(\tau) \) as the brownian motion with covariance matrix
\[ \Omega = 2\pi M^{-1}(\theta_\nu) f_{\omega\omega}(0) M^{-1}(\theta_\nu)', \]
the fractional brownian motion
\[ W(\tau; d-b) = \int_0^\tau \left( \tau - s \right)^{(d-b)-1} \frac{1}{\Gamma(d-b)} dW(s), \]
and then
\[ S = W(\tau; d-b) - \int_0^1 W(\tau; d-b) d\tau, \]
partitioning it in \( S_1, S_2 \) of dimensions \( r \times 1, (n - r) \times 1 \) respectively so that \( S = [S'_1, S'_2]' \). We also partition \( \Omega \) as
\[ \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \]
where \( \Omega_{11} \) is an \( r \times r \) matrix while \( \Omega_{22} \) is \( (n - r) \times (n - r) \), and finally introduce \( S_{1:2} \) as the brownian motion with covariance matrix \( \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \). Then,
\[ n^{(d-b)}(\hat{\nu}' - \nu') \to_d \int_0^1 dS_{1:2} S'_2 \left( \int_0^1 S_2 S'_2 \right)^{-1}. \]
Setting
\[ \gamma = (d, b, \varphi)' \]
and \( \Sigma(\gamma) \) as the matrix \( \Sigma \) after that the rows and columns corresponding to the elents of \( \nu \) have been removed,
\[ n^{1/2}(\hat{\gamma} - \gamma) \to_d N(0, \Sigma(\gamma)). \]

As already observed in Johansen (1991) (and by making use of exactly the same argument), if \( d - b > 1/2 \) the likelihood ratio test on \( \text{vec}(\nu) \), \( \gamma \) decomposes into a test for \( \text{vec}(\nu) \) and an independent test for \( \gamma \), so it is possible to conduct inference on \( \nu \) tacking \( \gamma \) as known and viceversa.
References


Figure 1.
Plot of the data for $i_1$ and $i_{12}$.

Figure 2.
Reaction of 1 month rate to an increase of 1 unit in the longer rates.
Figure 3.
Reaction of the rates to a 100 bp increase in the 1 month rate
### Table 1.
Semiparametric estimation of the memory parameters, tapered series
(optimal and user chosen bandwidths, local Whittle)

<table>
<thead>
<tr>
<th>Series:</th>
<th>m</th>
<th>a.s.e.</th>
<th>$d_1^{m}$</th>
<th>m</th>
<th>a.s.e.</th>
<th>$d_1^{m}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>i1</td>
<td>55</td>
<td>0.10</td>
<td>1.21</td>
<td>30</td>
<td>0.13</td>
<td>1.35</td>
</tr>
<tr>
<td>i3</td>
<td>127</td>
<td>0.06</td>
<td>1.16</td>
<td>30</td>
<td>0.13</td>
<td>1.34</td>
</tr>
<tr>
<td>i6</td>
<td>133</td>
<td>0.06</td>
<td>1.10</td>
<td>30</td>
<td>0.13</td>
<td>1.31</td>
</tr>
<tr>
<td>i12</td>
<td>133</td>
<td>0.06</td>
<td>1.11</td>
<td>30</td>
<td>0.13</td>
<td>1.24</td>
</tr>
</tbody>
</table>

### Table 2
Semiparametric estimation of the memory parameters, first difference of the series
(optimal and user chosen bandwidths, local Whittle)

<table>
<thead>
<tr>
<th>Series:</th>
<th>m</th>
<th>a.s.e.</th>
<th>$d_{\Delta i}$</th>
<th>m</th>
<th>a.s.e.</th>
<th>$d_{\Delta i}$</th>
<th>m</th>
<th>a.s.e.</th>
<th>$d_{\Delta i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>i1</td>
<td>41</td>
<td>0.08</td>
<td>-0.32</td>
<td>25</td>
<td>0.1</td>
<td>-0.14</td>
<td>48</td>
<td>0.07</td>
<td>-0.24</td>
</tr>
<tr>
<td>i3</td>
<td>45</td>
<td>0.07</td>
<td>-0.23</td>
<td>25</td>
<td>0.1</td>
<td>-0.11</td>
<td>48</td>
<td>0.07</td>
<td>-0.20</td>
</tr>
<tr>
<td>i6</td>
<td>54</td>
<td>0.07</td>
<td>-0.16</td>
<td>25</td>
<td>0.1</td>
<td>-0.08</td>
<td>48</td>
<td>0.07</td>
<td>-0.19</td>
</tr>
<tr>
<td>i12</td>
<td>69</td>
<td>0.06</td>
<td>-0.13</td>
<td>25</td>
<td>0.1</td>
<td>-0.05</td>
<td>48</td>
<td>0.07</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

### Table 3
Semiparametric tests for the equality of the orders
(optimal bandwidths, local Whittle estimates of the memory parameters, first differences)

<table>
<thead>
<tr>
<th>Series for</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>m</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$\hat{T}^2$</th>
<th>m</th>
<th>$\hat{T}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta i_1$</td>
<td>$\Delta i_3$</td>
<td>43</td>
<td>-0.30</td>
<td>-0.25</td>
<td>9.25</td>
<td>25</td>
<td>2.31</td>
<td></td>
</tr>
<tr>
<td>$\Delta i_1$</td>
<td>$\Delta i_6$</td>
<td>45</td>
<td>-0.27</td>
<td>-0.19</td>
<td>6.67</td>
<td>25</td>
<td>1.99</td>
<td></td>
</tr>
<tr>
<td>$\Delta i_1$</td>
<td>$\Delta i_{12}$</td>
<td>47</td>
<td>-0.25</td>
<td>-0.18</td>
<td>1.51</td>
<td>25</td>
<td>2.18</td>
<td></td>
</tr>
<tr>
<td>$\Delta i_3$</td>
<td>$\Delta i_6$</td>
<td>49</td>
<td>-0.20</td>
<td>-0.19</td>
<td>0.46</td>
<td>25</td>
<td>1.44</td>
<td></td>
</tr>
<tr>
<td>$\Delta i_3$</td>
<td>$\Delta i_{12}$</td>
<td>51</td>
<td>-0.18</td>
<td>-0.17</td>
<td>0.16</td>
<td>25</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>$\Delta i_6$</td>
<td>$\Delta i_{12}$</td>
<td>59</td>
<td>-0.17</td>
<td>-0.16</td>
<td>0.08</td>
<td>25</td>
<td>1.67</td>
<td></td>
</tr>
</tbody>
</table>

33
### Table 4
Marinucci and Robinson (2001) test of (no) cointegration

<table>
<thead>
<tr>
<th>Series for</th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(m)</th>
<th>(\bar{a}_e)</th>
<th>(\bar{a}_1)</th>
<th>(H_{m,d_1})</th>
<th>(\bar{a}_e)</th>
<th>(\bar{a}_2)</th>
<th>(H_{m,d_2})</th>
<th>(m)</th>
<th>(\bar{a}_e)</th>
<th>(H_{m,d_1})</th>
<th>(H_{m,d_2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i3)</td>
<td>22</td>
<td>-0.36</td>
<td>-0.20</td>
<td>4.47</td>
<td>22</td>
<td>-0.36</td>
<td>-0.17</td>
<td>6.34</td>
<td>25</td>
<td>-0.32</td>
<td>6.64</td>
<td>8.78</td>
</tr>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i6)</td>
<td>27</td>
<td>-0.32</td>
<td>-0.15</td>
<td>6.68</td>
<td>27</td>
<td>-0.32</td>
<td>-0.15</td>
<td>6.59</td>
<td>25</td>
<td>-0.31</td>
<td>6.21</td>
<td>10.57</td>
</tr>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i12)</td>
<td>21</td>
<td>-0.30</td>
<td>-0.22</td>
<td>0.90</td>
<td>21</td>
<td>-0.30</td>
<td>-0.19</td>
<td>1.80</td>
<td>25</td>
<td>-0.26</td>
<td>2.87</td>
<td>8.69</td>
</tr>
<tr>
<td>(\Delta i3)</td>
<td>(\Delta i6)</td>
<td>21</td>
<td>-0.33</td>
<td>-0.19</td>
<td>2.98</td>
<td>21</td>
<td>-0.33</td>
<td>-0.15</td>
<td>5.34</td>
<td>25</td>
<td>-0.31</td>
<td>8.19</td>
<td>10.46</td>
</tr>
<tr>
<td>(\Delta i3)</td>
<td>(\Delta i12)</td>
<td>21</td>
<td>-0.27</td>
<td>-0.15</td>
<td>2.31</td>
<td>21</td>
<td>-0.27</td>
<td>-0.11</td>
<td>4.11</td>
<td>25</td>
<td>-0.24</td>
<td>3.29</td>
<td>7.20</td>
</tr>
<tr>
<td>(\Delta i6)</td>
<td>(\Delta i12)</td>
<td>22</td>
<td>-0.22</td>
<td>-0.13</td>
<td>1.21</td>
<td>22</td>
<td>-0.22</td>
<td>-0.11</td>
<td>1.97</td>
<td>25</td>
<td>-0.20</td>
<td>2.51</td>
<td>4.32</td>
</tr>
</tbody>
</table>

### Table 5a.
Robinson and Yajima (2001) cointegration tests for \({i_1, i_3, i_6, i_{12}}\): four dimensional vector

<table>
<thead>
<tr>
<th>rank tested</th>
<th>(m = 48, m_1 = 52)</th>
<th>(m = 25, m_1 = 27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_0: {r = 3})</td>
<td>(\hat{\pi}_r)</td>
<td>(\hat{\pi}<em>r + z</em>{0.95}s_r/m^{1/2})</td>
</tr>
<tr>
<td>(H_0: {r = 2})</td>
<td>0.02052</td>
<td>0.02527</td>
</tr>
<tr>
<td>(H_0: {r = 1})</td>
<td>0.00085</td>
<td>0.00105</td>
</tr>
</tbody>
</table>

### Table 5b.
Robinson and Yajima (2001) cointegration tests for \({i_1, i_3, i_6, i_{12}}\):
pairwise test of \(H_0: \{r = 1\}\)

<table>
<thead>
<tr>
<th>Series for</th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(m)</th>
<th>(m_1)</th>
<th>(\hat{\pi}_1)</th>
<th>(\hat{\pi}<em>1 + z</em>{0.95}s_1/m^{1/2})</th>
<th>(m)</th>
<th>(m_1)</th>
<th>(\hat{\pi}_1)</th>
<th>(\hat{\pi}<em>1 + z</em>{0.95}s_1/m^{1/2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i3)</td>
<td>43</td>
<td>47</td>
<td>0.00260</td>
<td>0.00325</td>
<td>25</td>
<td>27</td>
<td>0.00140</td>
<td>0.00187</td>
<td></td>
</tr>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i6)</td>
<td>45</td>
<td>49</td>
<td>0.01188</td>
<td>0.01476</td>
<td>25</td>
<td>27</td>
<td>0.00616</td>
<td>0.00818</td>
<td></td>
</tr>
<tr>
<td>(\Delta i1)</td>
<td>(\Delta i12)</td>
<td>47</td>
<td>51</td>
<td>0.03201</td>
<td>0.03945</td>
<td>25</td>
<td>27</td>
<td>0.01599</td>
<td>0.02064</td>
<td></td>
</tr>
<tr>
<td>(\Delta i3)</td>
<td>(\Delta i6)</td>
<td>49</td>
<td>54</td>
<td>0.00487</td>
<td>0.00601</td>
<td>25</td>
<td>27</td>
<td>0.00195</td>
<td>0.00259</td>
<td></td>
</tr>
<tr>
<td>(\Delta i3)</td>
<td>(\Delta i12)</td>
<td>51</td>
<td>56</td>
<td>0.01858</td>
<td>0.02278</td>
<td>25</td>
<td>27</td>
<td>0.00700</td>
<td>0.09498</td>
<td></td>
</tr>
<tr>
<td>(\Delta i6)</td>
<td>(\Delta i12)</td>
<td>59</td>
<td>65</td>
<td>0.00482</td>
<td>0.00585</td>
<td>25</td>
<td>27</td>
<td>0.00286</td>
<td>0.00382</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.
Narrow Band FDLS estimates of the cointegrating parameter, and OLS.

<table>
<thead>
<tr>
<th>series</th>
<th>$\nu_{(m_2=5)}$</th>
<th>$\nu_{OLS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i3, i1$</td>
<td>1.008</td>
<td>1.005</td>
</tr>
<tr>
<td>$i6, i1$</td>
<td>0.998</td>
<td>0.990</td>
</tr>
<tr>
<td>$i12, i1$</td>
<td>0.945</td>
<td>0.931</td>
</tr>
</tbody>
</table>

Table 7.
Estimates of the memory parameters of the residuals, Narrow band FDLS and structural assumption ($\nu = 1$).

<table>
<thead>
<tr>
<th>series</th>
<th>$m$</th>
<th>a.s.e.</th>
<th>$b$</th>
<th>$m$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i3 - \nu_{(m_2=5)} i1$</td>
<td>59</td>
<td>0.07</td>
<td>0.18</td>
<td>25</td>
<td>0.22</td>
</tr>
<tr>
<td>$i6 - \nu_{(m_2=5)} i1$</td>
<td>49</td>
<td>0.07</td>
<td>0.21</td>
<td>25</td>
<td>0.30</td>
</tr>
<tr>
<td>$i12 - \nu_{(m_2=5)} i1$</td>
<td>48</td>
<td>0.07</td>
<td>0.33</td>
<td>25</td>
<td>0.44</td>
</tr>
<tr>
<td>$i3 - i1$</td>
<td>58</td>
<td>0.07</td>
<td>0.19</td>
<td>25</td>
<td>0.25</td>
</tr>
<tr>
<td>$i6 - i1$</td>
<td>49</td>
<td>0.07</td>
<td>0.21</td>
<td>25</td>
<td>0.30</td>
</tr>
<tr>
<td>$i12 - i1$</td>
<td>48</td>
<td>0.07</td>
<td>0.34</td>
<td>25</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 8a.
Breitung and Hassler (2002) cointegration tests for $\{i1, i3, i6, i12\}$: four dimensional vector

<table>
<thead>
<tr>
<th>rank tested</th>
<th>$\lambda_1$</th>
<th>$\sum_{j=1}^{n-r_0} \lambda_j$</th>
<th>$\chi^2$ $0.95(n-r_0)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0: {r = 3}$</td>
<td>0.01</td>
<td>0.01</td>
<td>3.84</td>
</tr>
<tr>
<td>$H_0: {r = 2}$</td>
<td>8.84</td>
<td>8.85</td>
<td>9.49</td>
</tr>
<tr>
<td>$H_0: {r = 1}$</td>
<td>29.21</td>
<td>38.05</td>
<td>16.92</td>
</tr>
<tr>
<td>$H_0: {r = 0}$</td>
<td>49.39</td>
<td>87.45</td>
<td>26.30</td>
</tr>
</tbody>
</table>

Table 8b.
Breitung and Hassler (2002) cointegration tests for $\{i1, i3, i6, i12\}$: pairwise tests

<table>
<thead>
<tr>
<th>series for $d_1$</th>
<th>$\Delta i1$</th>
<th>$\Delta i1$</th>
<th>$\Delta i1$</th>
<th>$\Delta i3$</th>
<th>$\Delta i3$</th>
<th>$\Delta i6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_2$</td>
<td>$\Delta i3$</td>
<td>$\Delta i6$</td>
<td>$\Delta i12$</td>
<td>$\Delta i6$</td>
<td>$\Delta i12$</td>
<td>$\Delta i12$</td>
</tr>
<tr>
<td>$\tilde{\lambda}_1$</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
<td>0.19</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>$\tilde{\lambda}_2$</td>
<td>22.82</td>
<td>28.61</td>
<td>22.36</td>
<td>22.64</td>
<td>17.67</td>
<td>14.61</td>
</tr>
</tbody>
</table>
Table 9.
Maximum Likelihood estimation of the cointegration relations

| Series | \( s_{3t} = (i3_t - 1.01i1_t) \) | \( s_{6t} = (i6_t - 1.01i1_t) \) | \( s_{12t} = (i12_t - 0.98i1_t) \) |

Table 10.
Maximum Likelihood estimation of the model for the residuals of the cointegration relations (the two matrices below correspond to \( B(1; \tilde{\varphi}) \) and \( B(2; \tilde{\varphi}) \) in (20)).

\[
\begin{align*}
\begin{bmatrix}
\Delta^{0.86}i_{11} \\
\Delta^{0.19}i_{33} \\
\Delta^{0.21}i_{66} \\
\Delta^{0.25}i_{12}
\end{bmatrix} &=
\begin{bmatrix}
0.52 & 1.08 & -0.57 & 0.41 \\
0.00 & -0.72 & 0.36 & 0.26 \\
-0.06 & -1.35 & 0.51 & 0.64 \\
-0.09 & -1.51 & 0.18 & 1.10
\end{bmatrix} +
\begin{bmatrix}
\Delta^{0.86}i_{11-1} \\
\Delta^{0.19}i_{33-1} \\
\Delta^{0.21}i_{66-1} \\
\Delta^{0.25}i_{12-1}
\end{bmatrix} +
\begin{bmatrix}
-0.17 & 0.15 & -0.95 & 0.45 \\
0.02 & -0.58 & 0.73 & -0.38 \\
0.06 & -0.19 & 0.45 & -0.40 \\
0.06 & 0.07 & 0.18 & -0.30
\end{bmatrix} +
\begin{bmatrix}
\tilde{\varepsilon}_{1t} \\
\tilde{\varepsilon}_{3t} \\
\tilde{\varepsilon}_{6t} \\
\tilde{\varepsilon}_{12t}
\end{bmatrix}
\end{align*}
\]

Table 11.
Estimate of the covariance matrix \( \Omega = \mathbb{E}(\varepsilon_t \varepsilon'_t) \) (as defined in equation 21).

<table>
<thead>
<tr>
<th>10000Ω</th>
<th>Correlation structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.025</td>
<td>0.03 0.04 -0.04 1.09 0.78 -0.34</td>
</tr>
<tr>
<td>0.03</td>
<td>0.06 0.08 -0.07 0.90 1.00 0.93 -0.42</td>
</tr>
<tr>
<td>0.04</td>
<td>0.08 0.11 -0.11 0.78 0.93 1.00 -0.50</td>
</tr>
<tr>
<td>-0.04</td>
<td>-0.07 -0.11 0.44 -0.34 -0.42 -0.50 1.00</td>
</tr>
</tbody>
</table>