Nonlinear Bivariate Comovements
of Asset Prices: Theory and Tests

M. Corazza\textsuperscript{1}, A.G. Malliaris\textsuperscript{2}, E. Scalco\textsuperscript{1}

\textsuperscript{1}Department of Applied Mathematics
University Ca’ Foscari of Venice (Italy)

\textsuperscript{2}Department of Economics
Loyola University of Chicago (Illinois, U.S.A.)

\textbf{Commodities Finance Centre – Commodities 2007}
January 17-19, 2007 - Birkbeck College, University of London (UK)
Outline

1) Introduction
2) A bit of literature
3) The three-steps methodology
4) An application to energy data
5) Cross-Greeks
6) Conclusions
1. Introduction (1)

The relevant importance of comovements among asset prices is due to several reasons:

1) its knowledge allows to use information on one asset for deducing the behaviour of the other one;

2) its presence in form of correlation is of interest to investors who wish to allocate their capitals in mean-variance portfolios;
1. Introduction (2)

3) its presence among globally traded assets may influence the coordination of economic policies;

4) scholars and policy makers are interested in it as an indication of the degree of financial integration.
2. A bit of literature (1)

The growing interest in this topic has resulted in a large volume of scientific contributions.

Some of those studies make use of:

- **spectral analysis** (see Granger et al. [1970]);
- **factor analysis** (see Ripley [1973]);
- **cluster analysis** (see Panton et al. [1976]).
2. A bit of literature (2)

A great deal of early research on the relationship between national equity markets mainly focuses on estimating *cross-country return correlations* and *covariances* (see, for instance, Among [1972] and Karolyi et al. [1996]).
2. A bit of literature (3)

In recent contributions, several authors follow methodologies based on:

- *autoregressive heteroskedastic (ARCH)* models (see Hamao *et al.* [1990]);
- *generalized ARCH models*;
- *multivariate cointegrations* (see Kasa [1992]);
- *structural vector autoregression (VAR) system* (see Eun *et al.* [1992]);
2. A bit of literature (4)

- vector error-correction models (see, Malliaris et al. [1996]);
- forecast error variance decomposition approaches;
- lag-augmented VAR systems (see Hamori et al. [2000]);
- Granger causality based tests (see Chen et al. [2004]).
2. A bit of literature (5)

Other methodologies that are worth mentioning are able to detect:
- the presence of *common trends* or the presence of *common features* (see Engle et al. (1993));
- the presence of *co-dependence* (see Broome *et al.* [2000]).
3. Our three-steps methodology (1)

Our approach allows
- the *evaluation* and
- the *statistical testing*

of non-linearly driven comovements between two random variables.

Moreover, when such dependence relationship is detected, our approach provides also a *polynomial approximation* of it.
3. Our three-steps methodology (2)

**Step 1** - We specify a simple index able to evaluate any intermediate degree of the nonlinear bivariate dependence.

**Step 2** - We propose a procedure to test the statistical meaningfulness of the index itself.

**Step 3** - We propose an algorithm to provide a polynomial approximation of the unknown bivariate dependence relationship.
3. Our three-steps methodology (3)

The simple index

Consider two (discrete) time series:
\[ \{X_1(t), t = t_1, \ldots, t_N\} \text{ and } \{X_2(t), t = t_1, \ldots, t_N\}. \]

The simple index we propose for evaluating the bivariate dependence between \( X_1(t) \) and \( X_2(t) \) is defined as follows:
3. Our three-steps methodology (4)

\[ \delta_{1,2} = \sum_{t=t_0}^{t_N} \frac{\Delta(t)_{1,2}}{N-1}, \]

where

\[ \Delta(t)_{1,2} = \begin{cases} 
-1 & \text{if } [X_1(t) - X_1(t-1)] \cdot [X_2(t) - X_2(t-1)] < 0 \\
1 & \text{if } [X_1(t) - X_1(t-1)] \cdot [X_2(t) - X_2(t-1)] \geq 0 
\end{cases} \]
3. Our three-steps methodology (5)

Notice that:

- \( \delta_{1,2} \in [-1,1] \) (property of normalization of the first type);
- \( \delta_{1,2} \) is defined for every pair of (discrete) time series (property of existence);
- \( \delta_{1,2} = \delta_{2,1} \) (property of symmetry).

So, our index is a **scalar measure of dependence** in the sense of Szegö [2005].
3. Our three-steps methodology (6)

**Proposition 1** - Let \( f(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) be the bivariate dependence relationship between \( X_1(t) \) and \( X_2(t) \), i.e. \( X_1(t) = f(X_2(t)) + \varepsilon(t) \), and let \( f(\cdot) \) be infinite times derivable in \( m_2 = E(X_2(t)) \). If

\[
\frac{f^{(i)}(m_2)}{i!} \binom{i}{i-j} (-m_2)^{i-j} = 0, \quad \forall j = 2, \ldots, +\infty
\]

where \( f^{(i)}(\cdot) \) indicates the \( i \)-derivatives of \( f(\cdot) \), then the bivariate dependence relationship is affine.
3. Our three-steps methodology (7)

The testing procedure

**Step 1** - We define the r.v. $\delta_{s;1,2}$ as the index $\delta_{1,2}$, but applied to the two time series once both have been shuffled.

**Step 2** - We defined the quantity $\Delta \delta = \delta_{1,2} - \delta_{S;1,2}$ and generate the series $\{\Delta \delta(j), J = 1, \ldots, M\}$ by shuffling for $M$ times the two time series.
3. Our three-steps methodology (8)

**Step 3** - We determine the *sample mean* and the *sample standard deviation* of $\Delta\delta$ as follows:

$$m_{\Delta\delta} = \frac{1}{M} \sum_{j=1}^{M} (\Delta\delta(j))$$

and

$$s_{\Delta\delta} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (\Delta\delta(j) - m_{\Delta\delta})^2}.$$
3. Our three-steps methodology (9)

**Step 4** - For $M$ large enough we perform the following (bilateral) $t$-test:

$$
\begin{aligned}
H_0 : & \quad m_{\Delta \delta} = 0, \text{ i.e. } X_1(t) \text{ and } X_2(t) \text{ are independent} \\
H_1 : & \quad m_{\Delta \delta} \neq 0, \text{ i.e. } X_1(t) \text{ and } X_2(t) \text{ are dependent}
\end{aligned}
$$

**Step 5** - We recall from basic statistic that

$$
\frac{m_{\Delta \delta} - \Delta \delta}{s_{\Delta \delta} / \sqrt{M - 1}} \rightarrow N(0,1) \text{ as } M \rightarrow +\infty
$$
3. Our three-steps methodology (10)

**Step 6** – Finally, if $H_0$ is rejected, then we perform two more unilateral $t$-tests in order to verify whether the $\delta_{1,2}$-dependence between $X_1(t)$ and $X_2(t)$ is positive or negative.
3. Our three-steps methodology (11)

The polynomial approximation

We analytically model the unknown bivariate $\delta_{1,2}$-dependence relationship $X_1(t) = f(X_2(t)) + \varepsilon(t)$.

In particular, we look for a polynomial approximation of the type

$$f(X_2(t)) = \sum_{j=0}^{J} a_j X_2^j(t) + r(J + 1)$$

(1)
3. Our three-steps methodology (12)

where

\[ J \] is the truncation order of the Taylor’s series;

\[ a_j = \sum_{i=j}^J \frac{f^{(i)}(m_2)}{i!} \binom{i}{i-j} (-m_2)^{i-j}; \]

\( r(J+1) \) is a suitable remainder function.

- Notice that \( J \) plays a crucial role.
3. Our three-steps methodology (13)

- How to detect the “optimal” value of $J$?

**Step 1** – We let $D = \{(X_1(t), X_2(t)), t = t_1, \ldots, t_N\}$ as the starting data set.

**Step 2** – We split $D$ into a learning subset $D_L$ and a validation subset $D_V$ (such that $D_L \cup D_V = D$ and $D_L \cap D_V = \emptyset$);
3. Our three-steps methodology (14)

**Step 3** – We consider a finite series of polynomials of form (1) with \( J = 0, \ldots, \bar{J} \), where \( \bar{J} \) is a pre-established integer value.

**Step 4** – For each of the polynomials we estimate the parameters \( a_0, \ldots, a_J \) via OLS by using \( D_L \), and evaluate the index \( \delta_{1,2} \)

between \( \hat{X}_1(t) = \sum_{j=0}^{J} \hat{a}_j X_2^j(t) \) and \( X_2(t) \) by using \( D_V \).
3. Our three-steps methodology (15)

Step 5 – Finally, we choose as “best” approximating polynomial the one associated with the highest absolute value of $\delta_{1,2}$.
4. An application to energy data (1)

**Step 1** – We start by considering the bivariate time serie \( \{(X_1(t), X_2(t)), t = t_1, ..., t_N\} \).

**Step 2** – From it we split the chronologically last 10% in order to utilize them as forecasting data set \( D_F \); we use the remaining 90% as the starting data set, and we split it into \( D_L \) and \( D_V \);

**Step 3** – Finally we apply our methodology.
4. An application to energy data (2)

We utilize (discrete) time series constituted by 2,026 daily spot closing prices (from January 3, 1994 to December 31, 2002) of three energy assets:

- **Heating Oil:** \(\{X_1(t), t = t_1, \ldots, t_{2,026}\}\);
- **Gasoline:** \(\{X_2(t), t = t_1, \ldots, t_{2,026}\}\);
- **Crude Oil:** \(\{X_3(t), t = t_1, \ldots, t_{2,026}\}\).
4. An application to energy data (3)

<table>
<thead>
<tr>
<th>Random variables</th>
<th>$\delta_{i,j}$</th>
<th>Bilateral t test</th>
<th>Check</th>
<th>$\rho_{i,j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{CO}(t),\ X_G(t)$</td>
<td>0.35407</td>
<td>$R$</td>
<td>$P$</td>
<td>0.40761</td>
</tr>
<tr>
<td>$X_{CO}(t),\ X_{HO}(t)$</td>
<td>0.39259</td>
<td>$R$</td>
<td>$P$</td>
<td>0.49802</td>
</tr>
<tr>
<td>$X_G(t),\ X_{HO}(t)$</td>
<td>0.48642</td>
<td>$R$</td>
<td>$P$</td>
<td>0.58000</td>
</tr>
</tbody>
</table>
4. An application to energy data (4)

<table>
<thead>
<tr>
<th>R.v.s</th>
<th>Polynomial approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{CO}(t), X_G(t)$</td>
<td>$X_{CO}(t) = 1.68731 + 31.36198 \times X_G(t)$</td>
</tr>
<tr>
<td>$X_{CO}(t), X_{HO}(t)$</td>
<td>$X_{CO}(t) = -9.38380 + 97.98516 \times X_{HO}(t) - $</td>
</tr>
<tr>
<td></td>
<td>$- 116.50908 \times X_{HO}^2(t) + 63.03937 \times X_{HO}^3(t)$</td>
</tr>
<tr>
<td>$X_G(t), X_{CO}(t)$</td>
<td>$X_G(t) = 0.03803 + 0.02687 \times X_{CO}(t)$</td>
</tr>
<tr>
<td>$X_G(t), X_{HO}(t)$</td>
<td>$X_G(t) = 0.12943 + 0.79407 \times X_{HO}(t)$</td>
</tr>
<tr>
<td>$X_{HO}(t), X_{CO}(t)$</td>
<td>$X_{HO}(t) = -3.80625 + 0.87743 \times X_{CO}(t) - 0.06879 \times X_{CO}^2(t) +$</td>
</tr>
<tr>
<td></td>
<td>$+ 0.00240 \times X_{CO}^3(t) - 0.00003 \times X_{CO}^4(t)$</td>
</tr>
<tr>
<td>$X_{HO}(t), X_G(t)$</td>
<td>$X_{HO}(t) = -0.19987 + 2.37713 \times X_G(t) - 2.98411 \times X_G^2(t) +$</td>
</tr>
<tr>
<td></td>
<td>$+ 1.89758 \times X_G^3(t)$</td>
</tr>
</tbody>
</table>
$X_{H0}(t)$ (continuous line) and $X_{H0}(t) = \rho(X_{C0}(t))$ (dotted line)
$X_{H0}(t)$ (continuous line) and $X_{H0}(t) = p(X_G(t))$ (dotted line)
$X_{C_0}(t)$ (continuous line) and $X_{C_0}(t) = \rho(X_C(t))$ (dotted line)
$X_C(t)$ (continuous line) and $X_{C0}(t) = p(X_H(t))$ (dotted line)
$X_0(t)$ (continuous line) and $X_0(t) = p(X_{CO}(t))$ (dotted line)
$X_0(t)$ (continuous line) and $X_0(t) = p(X_{H0}(t))$ (dotted line)
5. Cross-Greeks (1)

We present some results concerning an utilization of the polynomial approximation of the bivariate dependence relationship in the research field of option contracts: the \textit{cross-hedging} one.
5. Cross-Greeks (2)

**Proposition 2** - Let the usual hypotheses concerning the B&S environment hold, and let $X_1(t)$ and $X_2(t)$ be the prices of two assets, both defined on a given interval $[t_0, t_1]$. If

$$X_1(X_2(t)) = \sum_{i=0}^{K} a_i X_2^i(t)$$

then
5. Cross-Greeks (3)

- Cross-delta_{call} = \Phi(d_1^*) X_1^{(1)} (X_2 (t))

- Cross-vega_{call} = X_1 (X_2 (t)) \sqrt{\tau} \Phi^{(1)} (d_1^*)

- Cross-theta_{call} = \frac{X_1(X_2 (t))\sigma}{2\sqrt{\tau}} \Phi^{(1)} (d_1^*) + Xre^{-r\tau} \Phi(d_2^*)

- Cross-rho_{call} = X \tau e^{-r\tau} \Phi(d_2^*)

- Cross-gamma_{call} = 
  \frac{\Phi^{(1)} (d_1^*)}{X_1(X_2 (t))\sigma\sqrt{\tau}} \left[ X_1^{(1)} (X_2 (t)) \right]^2 + \Phi(d_1^*)X_1^{(2)} (X_2 (t))
5. Cross-Greeks (4)

- Cross-delta_{put} = \left[ \Phi \left( d_1^* \right) - 1 \right] X_1^{(1)} \left( X_2 \left( t \right) \right)

- Cross-vega_{put} = X_1 \left( X_2 \left( t \right) \right) \sqrt{\tau} \Phi^{(1)} \left( d_1^* \right)

- Cross-theta_{put} = \frac{X_1(X_2(t))\sigma}{2\sqrt{\tau}} \Phi^{(1)}(d_1^*) + Xr e^{-r\tau} \left[ \Phi(d_2^*) - 1 \right]

- Cross-rho_{put} = Xr e^{-r\tau} \left[ \Phi \left( d_2^* \right) - 1 \right]

- Cross-gamma_{put} = \frac{\Phi^{(1)}(d_1^*)}{X_1(X_2(t))\sigma\sqrt{\tau}} \left[ X_1^{(1)}(X_2(t)) \right]^2 + \left[ \Phi(d_1^*) - 1 \right] X_1^{(2)}(X_2(t))
5. Cross-Greeks (5)

where

\[ d_1^* = \frac{\log(X_1(X_2(t))/X) + r\tau + \sigma^2 \tau / 2}{\sigma \sqrt{\tau}}; \]

\[ d_2^* = d_1^* - \sigma \sqrt{\tau}. \]
5. Cross-Greeks (6)

**Example** – We determine the analytical approximations of the Greeks for the European call and put options in terms of the price of *heating oil* when the underlying is the price of *gasoline*.

We recall that:

\[ \hat{X}_G(t) = 0.12943 + 0.79407 X_{HO}(t) \]
5. Cross-Greeks (7)

The cross-Greeks for the European call options are:

\[ \text{cross-delta}_{\text{call}} = 0.79407 \cdot \Phi(d_1^*) \]

\[ \text{cross-vega}_{\text{call}} = \left[0.12943 + 0.79407 X_{HO}(t)\right] \cdot \sqrt{\tau} \Phi^{(1)}(d_1^*) \]

\[ \text{cross-theta}_{\text{call}} = \]
\[ = \left(1/2\sqrt{\tau}\right) \left[0.12943 + 0.79407 X_{HO}(t)\right] \Phi^{(1)}(d_1^*) + Xre^{-r\tau} \Phi(d_2^*) \]

\[ \text{cross-rho}_{\text{call}} = Xre^{-r\tau} \Phi(d_2^*) \]

\[ \text{cross-gamma}_{\text{call}} = \frac{0.63054 \Phi^{(1)}(d_1^*)}{\left[0.12943 + 0.79407 X_{HO}(t)\right] \sigma \sqrt{\tau}} \]
5. Cross-Greeks (8)

The cross-Greeks for the European put options are:

\[
\text{cross-delta}_{\text{put}} = 0.79407 \cdot \left[ \Phi(d_1^*) - 1 \right]
\]

\[
\text{cross-vega}_{\text{put}} = \left[ 0.12943 + 0.79407 X_{HO}(t) \right] \sqrt{\tau} \Phi^{(1)}(d_1^*)
\]

\[
\text{cross-theta}_{\text{put}} = (1/2\sqrt{\tau}) \left[ 0.12943 + 0.79407 X_{HO}(t) \right] \sigma \cdot \Phi^{(1)}(d_1^*) + X e^{-r\tau} \left[ \Phi(d_2^*) - 1 \right]
\]

\[
\text{cross-rho}_{\text{put}} = X \tau e^{-r\tau} \left[ \Phi(d_2^*) - 1 \right]
\]

\[
\text{cross-gamma}_{\text{put}} = \frac{0.63054 \cdot \Phi^{(1)}(d_1^*)}{\left[ 0.12943 + 0.79407 X_{HO}(t) \right] \sigma \sqrt{\tau}}
\]
6. Concluding remarks

Checking the validation approach by means of further applications of the three-steps methodology.

Developing in a more formal way the visual out-of-sample check.

Providing generalizations of our approach (i.e. time lagged relationships, multivariate no-lagged and lagged relationships, …).