Macro Factors in Oil Futures Returns

Yannick LE PEN
PSL*, Université de Paris-Dauphine, LEDa-CGEMP

Benoît SÉVI
Aix-Marseille Université, CNRS & EHESS

2ème Séminaire Académique d’Economie de l’Energie
The framework

• Which causes affect oil price in the late 2000s?

• Speculation or rising demand?

• Büyüksahin et al. (2008), Hamilton (2009), Kilian (2009), Büyüksahin and Harris (2011), Parsons (2010), Kaufmann (2011) and Tang and Xiong (2011):
  1. No evidence of causality from speculation to price.

• Hamilton (2009): the 2008 oil price increase attributed to a “demand shock” which may have its origin in Asia and more particularly in China.
The framework (con’t)

- Kilian and Vega (2011)
  - No evidence of an impact of US macroeconomic news on daily price changes in the oil spot market.
  - But:
    - Macroeconomic news may impact longer maturity futures contracts
    - U.S. news: a part of the story.
- We extend Kilian and Vega (2011) analysis and consider a set of macroeconomic variables representative of developed and emerging countries.
Two objectives of the paper

1. How useful is a large set of international real and nominal variables in explaining crude oil return?
   - We gather a set of 187 real and nominal macroeconomics variables from developed and emerging countries.
   - We apply “Large approximate factor model” (Stock and Watson (2002) to extract factors from these data.
   - These factors represent demand related “fundamentals”.
   - Avoid to select an a priori set of explanatory variables. We expect to minimize the risk of omitted variable.

2. How can we interpret the factors that have the best explanatory power?
   - We look at the explanatory power of each factor for the original series.
   - A criterion proposed by Ludvigson and Ng (2009)
Related literature

- Applies the large factor model method.
- Uses variables related to the US economy and oil and oil derivates times series. Criticized by Alquist et al. (2011) for doing so.
- We include in our database real and nominal variables from developed and emerging countries.
Sketch of results

- Our “best” model explains around 38% of oil returns variability.
- The factor with the highest explanatory power is mainly correlated with real variables from emerging countries.
Outline of the presentation

Introduction

Data

Factor computation

Fitting oil return

Conclusion
Oil futures and Macroeconomic database

- Monthly futures prices for the NYMEX WTI
- Time period: 1993:11 to 2010:03 (197 monthly observations).
- We use monthly observations to match with macroeconomic variables frequency.
Oil futures price
Oil futures return

Return is computed as the price log difference
# Descriptive statistics for monthly crude oil returns

<table>
<thead>
<tr>
<th></th>
<th>$r_{oil,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0077</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3045</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.4340</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0991</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.5770</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.6766</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>33.83**</td>
</tr>
<tr>
<td>Nb of Obs</td>
<td>196</td>
</tr>
</tbody>
</table>

*Note: “**” denotes a rejection of the null hypothesis of a Gaussian distribution at the 5% level.*

- Negative skewness and excess kurtosis: non Gaussian distribution.
**Macroeconomic database**

- 187 international macroeconomic and nominal variables representative of the world economy.
- 128 variables for Developed economies and 59 variables for emerging countries.
- 103 real variables (73 for developed countries, 30 for emerging countries) and 84 nominal variables (55 for developed and 29 for emerging countries).
- Differ from Stock and Watson (2005) and Ludvigson and Ng (2009) mainly focused on the US economy.
- Before computation, data are stationarized using the appropriate transformation if needed (first difference, log first difference,...).
- All data extracted from DataStream.
Large approximate factor model

• Let $x_{i,t} =$ observation of the $i^{th}$ time series ($i = 1, ..., N$) at date $t$ ($t = 1, ..., T$)

• Selecting relevant variables among $N$ variables when $N$ is large is not possible; we then resort to a set of $r$ factors:

$$x_{it} = \lambda_i' F_t + e_{it}$$

• $F_t$: vector of the $r$ common factors.
• $e_{it}$: idiosyncratic error
• $\lambda_i$: factor loadings of the (static) common factors
• Computation of factors via principal component analysis.
Estimating factors

- Assumption of $k$ factors
- $T \times k$ matrix of factors $F^k$ and corresponding $N \times T$ loading matrix $\Lambda^k$ estimated through the principal component method.
- These estimates solve the following optimization problem:

$$
\min S(k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \lambda_i^k F_t^k)^2
$$

subject to the normalization $\Lambda^k' \Lambda^k / N = I_k$.

- $X$ as the $T \times N$ matrix of observations
- $\hat{\Lambda}^k$ equal to the eigenvectors of the largest $k$ eigenvalues of $X'X$
- $\hat{F}^k = N^{-1} X' \hat{\Lambda}^k$
Selecting the number of factors

- Bai and Ng (2002) information criteria: an extension to factor model of usual information criteria (AIC..).

\[
PCP_i(k) = \hat{S}(k) + k\bar{\sigma}^2 g_i(N, T)
\]

\[
IC_i(k) = \ln(\hat{S}(k)) + kg_i(N, T)
\]

- \(\hat{S}(k)\) residual sum of square, \(g_i\) penalty function, \(\bar{\sigma}^2 = \hat{S}(k_{max})\) for a pre-specified value \(k_{max}\)

- Kapetanios (2009) sequential test for determining the number of factors
### Selecting the number of factors

<table>
<thead>
<tr>
<th>Method</th>
<th>No of static factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MED</td>
<td>2</td>
</tr>
<tr>
<td>$IC_1$</td>
<td>3</td>
</tr>
<tr>
<td>$IC_2$</td>
<td>2</td>
</tr>
<tr>
<td>$IC_3$</td>
<td>20</td>
</tr>
<tr>
<td>$IC_4$</td>
<td>20</td>
</tr>
<tr>
<td>$PCP_1$</td>
<td>9</td>
</tr>
<tr>
<td>$PCP_2$</td>
<td>7</td>
</tr>
<tr>
<td>$PCP_3$</td>
<td>20</td>
</tr>
<tr>
<td>$PCP_4$</td>
<td>20</td>
</tr>
</tbody>
</table>

**Notes**: MED denotes the number of factors given by the Maximum eigenvalue algorithm. $IC_i$ and $PCP_i$ respectively denote the number of factors given by the information criteria $IC$ and $PCP$ estimated with penalty function $g_i(N, T)$.

- No agreement on the estimated number of factors (a problem often encountered).
Selecting the number of factors: summary statistics

<table>
<thead>
<tr>
<th>( \hat{F}_{t,i} )</th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>( \rho_3 )</th>
<th>( R_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1614</td>
<td>0.1256</td>
<td>0.3176</td>
<td>0.0975</td>
</tr>
<tr>
<td>2</td>
<td>0.1357</td>
<td>0.0805</td>
<td>0.3110</td>
<td>0.1619</td>
</tr>
<tr>
<td>3</td>
<td>-0.0748</td>
<td>0.0145</td>
<td>-0.0294</td>
<td>0.2030</td>
</tr>
<tr>
<td>4</td>
<td>-0.0765</td>
<td>-0.0910</td>
<td>0.1508</td>
<td>0.2355</td>
</tr>
<tr>
<td>5</td>
<td>-0.2180</td>
<td>-0.0763</td>
<td>0.1213</td>
<td>0.2654</td>
</tr>
<tr>
<td>6</td>
<td>0.1801</td>
<td>0.0388</td>
<td>0.0267</td>
<td>0.2927</td>
</tr>
<tr>
<td>7</td>
<td>0.0721</td>
<td>0.2765</td>
<td>0.2744</td>
<td>0.3185</td>
</tr>
<tr>
<td>8</td>
<td>0.4086</td>
<td>0.5013</td>
<td>0.3332</td>
<td>0.3418</td>
</tr>
<tr>
<td>9</td>
<td>-0.0066</td>
<td>-0.0305</td>
<td>-0.0379</td>
<td>0.3636</td>
</tr>
</tbody>
</table>

Note: \( \rho_i \) denotes the \( i^{th} \) autocorrelation. \( R_i^2 \): fraction of total variance in the data explained by factors 1 to \( i \).

- We select the first 9 factors which explain 36 \% of the total variance in the data.
Selecting a model for oil futures returns

- Factors are selected according to their individual explanatory power.
- \( \hat{F}_{1,t} \) has the highest explanatory power (around 14%).
- \( \hat{F}_{3,t} \) and \( \hat{F}_{9,t} \) are excluded as they have almost none explanatory power.
- We consider all linear regression for all subsets of the 7 remaining factors.
- We select the regression which minimizes the BIC criterion.
- Selected linear regression:

\[
    r_{oil,t} = \alpha + \beta \hat{F}_t + u_t = \alpha_1 + \beta_1 \hat{F}_{1,t} + \beta_2 \hat{F}_{2,t} + \beta_4 \hat{F}_{4,t} + \beta_7 \hat{F}_{7,t} + u_t
\]
# OLS estimate of the selected regression

<table>
<thead>
<tr>
<th>$r_{oil,t}$</th>
<th>Intercept 0.0077</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.38)</td>
</tr>
<tr>
<td>$\widehat{F}_1$</td>
<td>-0.1217***</td>
</tr>
<tr>
<td></td>
<td>(-7.49)</td>
</tr>
<tr>
<td>$\widehat{F}_2$</td>
<td>-0.1489***</td>
</tr>
<tr>
<td></td>
<td>(-7.95)</td>
</tr>
<tr>
<td>$\widehat{F}_4$</td>
<td>0.0957***</td>
</tr>
<tr>
<td></td>
<td>(3.07)</td>
</tr>
<tr>
<td>$\widehat{F}_7$</td>
<td>0.1454***</td>
</tr>
<tr>
<td></td>
<td>(4.13)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.3787</td>
</tr>
<tr>
<td>$\overline{R}^2$</td>
<td>0.3657</td>
</tr>
</tbody>
</table>

Notes: (i) t-statistics are reported in parenthesis under the estimates. (ii) For each test ***, **, and * respectively denotes rejection of the null hypothesis of insignificant coefficient at the 1%, 5% and 10% levels.
Interpreting factors

- Ludvigson and Ng (2009) suggest a simple method to interpret the estimated factors.
- Each original variable is regressed on a single factor to measure the correlation between the former and the latter.
- The $R^2$ are reported on a graph with a given order.
- The factor is considered as representative of the variables with highest $R^2$.
- Our 187 series classified into four categories according to the characteristics real variable/nominal variable and developed countries/emerging countries.
• $\hat{F}_1$ interpreted as a real factor.
• Mostly correlated with real variables from emerging countries
• An evidence of the growing weight of emerging countries in shaping oil price.
• Highest $R^2$ for nominal variables of developed countries
• Can be interpreted as a “nominal” factor.
Interpreting factor $\hat{F}_2$

- More difficult to interpret
- Highest $R^2$ with a subset of aggregate consumption of developed countries
Interpreting factor $\hat{F}_7$

Also correlated with real variables of developed countries
Limits and extensions

1. Enlarging the database
   • Data on inventories and production
   • Checking the relevance of the series in the database (Boivin and Ng (2006))

2. More sophisticated econometric methods
   • Comparison with dynamic factor models (more appropriate for a forecasting exercise)
   • Bootstrapping factors (Ludvigson and Ng (2009, 2010) and Gospodinov and Ng (2010)) because factors are estimated quantities.
   • Using times series of different frequencies (MIDAS).
Limits and extensions

3. The role of speculation

- Bunn, Chevallier, Le Pen and Sevi (2013) “Fundamental and Financial Influences on the Comovement of Oil and Gas price”,
- Le Pen and Sevi (2013) “Futures trading and the excess comovement of commodity prices”.
- We show that the commodity returns correlation is highly related to:
  - the Han (2008) index of speculative activity,
  - the De Roon et al. (2000) index of hedging pressure,
- We conclude that trading activity on these markets has an impact on intercommodity correlation.