Macro Factors in Oil Futures Returns

Yannick LE PEN
PSL*, Université de Paris-Dauphine, LEDa-CGEMP
Benoît SÉVI
Aix-Marseille Université, CNRS & EHESS

$2^{ème}$ Séminaire Académique d'Economie de l'Energie

The framework

- Which causes affect oil price in the late 2000s?
- Speculation or rising demand?
- Büyükşahin *et al.* (2008), Hamilton (2009), Kilian (2009), Büyükşahin and Harris (2011), Parsons (2010), Kaufmann (2011) and Tang and Xiong (2011) :
    1. No evidence of causality from speculation to price.
    2. Minor role of trading activity in the NYMEX WTI in the 2008 price peak formation.
- 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)

- Standard analysis of macro factors for crude oil returns: Brown and Yücel (2002) or Lescaroux and Mignon (2008) (among others).
- 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 developped 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 developped 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

- Zagaglia (2010): a paper related to ours.
- 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

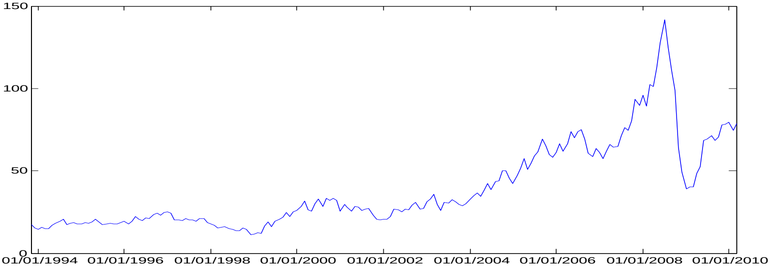
- Our "best" model explains around 38% of oil returns variability
- The factor with the highest explanatory power is mainly correlated with real variables from emerging countries.

Outline of the presentation

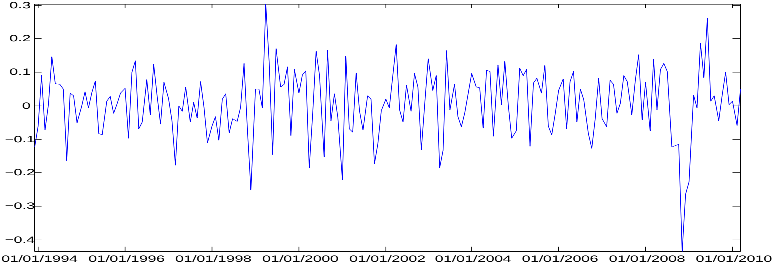
Oil futures and Macroeconomic database

- 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

	$r_{oil,t}$
Mean	0.0077
Maximum	0.3045
Minimum	-0.4340
Std. Dev.	0.0991
Skewness	-0.5770
Kurtosis	4.6766
Jarque-Bera	33.83**
Nb of Obs	196

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
  - λ_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 Λ^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..).

$$
\begin{aligned}
PCP_i(k) &= \hat{S}(k) + k\bar{\sigma}^2 g_i(N,T) \\
IC_i(k) &= \ln(\hat{S}(k)) + k g_i(N,T)
\end{aligned}
$$

- $\hat{S}(k)$ 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

Method	No of static factors
MED	2
IC_1	3
IC_2	2
IC_3	20
IC_4	20
PCP_1	9
PCP_2	7
PCP_3	20
PCP_4	20

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

$\widehat{F}_{t,i}$ i	ρ_1	ρ_2	ρ_3	R_i^2
1	0.1614	0.1256	0.3176	0.0975
2	0.1357	0.0805	0.3110	0.1619
3	-0.0748	0.0145	-0.0294	0.2030
4	-0.0765	-0.0910	0.1508	0.2355
5	-0.2180	-0.0763	0.1213	0.2654
6	0.1801	0.0388	0.0267	0.2927
7	0.0721	0.2765	0.2744	0.3185
8	0.4086	0.5013	0.3332	0.3418
9	-0.0066	-0.0305	-0.0379	0.3636

Note: ρ_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:

$$r_{oil,t} = \alpha + \beta \hat{F}_t + u_t = \alpha_1 + \beta_1 \hat{F}_t^1 + \beta_2 \hat{F}_t^2 + \beta_4 \hat{F}_t^4 + \beta_7 \hat{F}_t^7 + u_t$$

OLS estimate of the selected regression

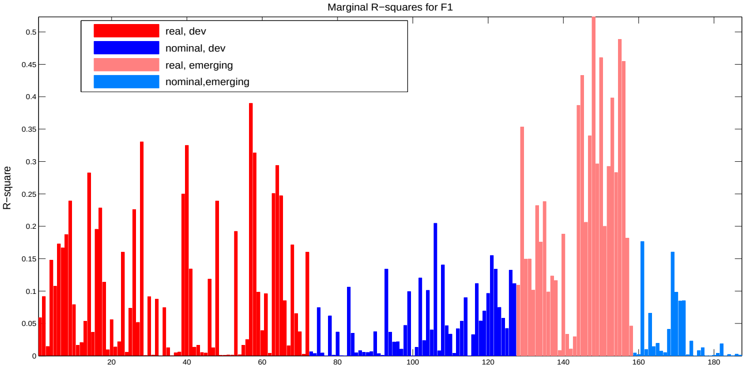
	$r_{oil,t}$
Intercept	0.0077
	(1.38)
$\widehat{F}_1$	-0.1217***
	(-7.49)
$\widehat{F}_2$	-0.1489***
	(-7.95)
$\widehat{F}_4$	0.0957***
	(3.07)
$\widehat{F}_7$	0.1454***
	(4.13)
R^2	0.3787
$\overline{R}^2$	0.3657

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

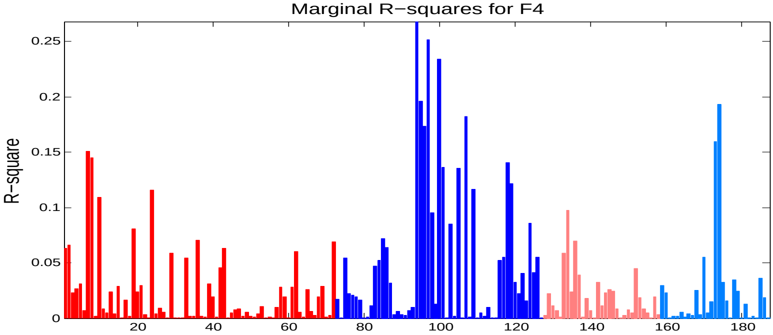
- Ludvigson and Ng (2009) suggest a simple method to interpret the estimated factors.
- Each original variable is regressed on a single factor to measure the correlation between the former and the latter.
- The R^2 are reported on a graph with a given order.
- The factor is considered as representative 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.

Interpreting factor $\hat{F}_1$



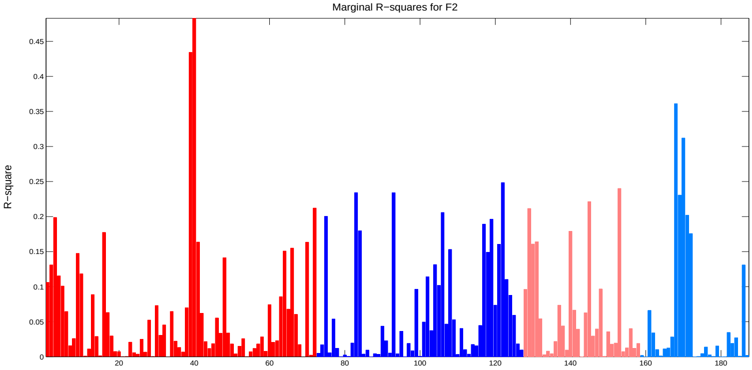
- $\hat{F}_t^1$ interpreted as a real factor.
- Mostly correlated with real variables from emerging countries
- An evidence of the growing weight of emerging countries in shaping oil price.

Interpreting factor $\hat{F}_4$



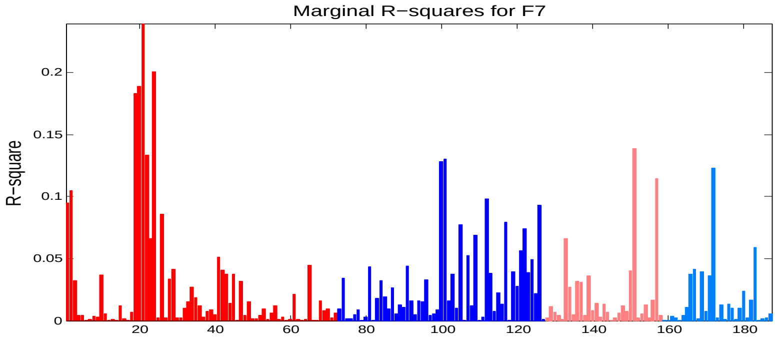
- Highest R^2 for nominal variables of developed countries
- Can be interpreted as a "nominal" factor.

Interpreting factor $\hat{F}_2$



- More difficult to interpret
- Highest R^2 with a subset of agregate 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,

     on commodity markets.
   - We conclude that trading activity on these markets has an impact on intercommodity correlation.