Big Fish: Oil Markets and Speculation

Elisa Scarpa
Head of Market Analysis, Forecasting & Structuring
Edison Trading S.p.A.

Institut Henri Poincaré
Paris, 9th October 2015

EDISON
Market Analysis and Forecasting

Big Fish

“Did you ever think maybe you're not too big—but, maybe this town's just too small?”

“Most *men*, they'll *tell* you a *story* straight true. It won't be *complicated*, but it won't be interesting either.”

Outline

- Motivation
- Theory and empirical findings
- Speculation indexes: a matter of fundamentals and economics
- The Empirical Analysis
  - Data
  - The equilibrium price
  - Hedging needs
  - The model and the econometric techniques
  - Empirical results
- Main empirical findings

Motivation

- There has been a strong increase in the number of financial operators taking part in the oil future markets *(‘financialization’* of commodity markets).
- At the same time, we observe an evolution in hedging strategies of commercial operators together with an increasing trend in the size and complexity of the physical oil market.
- Several authors have associated oil price dynamics to ‘speculators’.
- We believe that few important variables have not been considered in the empirical literature and that **‘speculators’ indeed act as oil price *stabilizers.***
- Furthermore, level of prices far from the equilibrium target cannot be explained by financial positions in the market.

Theory and empirical findings

- The market is characterized mainly by two agents:
  - **<u>producers and commercial operators</u>** with the need to hedge their positions (‘hedging needs’) and
  - **<u>speculators</u>** with the need to make money (‘money needs’).
- The spot price approaches the equilibrium price, given a stable supply demand balance, when *hedging needs* are met via *money needs*.
- We use a conditional ECM (applying Pesaran et al. (2001)’s bounds tests) to examine hedging needs and to detect the effects of money needs on deviations from the equilibrium price.
- Fundamentals variables play together with the term structure a key role in ensuring the equilibrium of crude oil future markets.

Existing literature

On the *financialization* of crude oil markets

- Speculative positions taken by institutional investors had resulted in increases in futures and spot commodity prices (e.g. Masters (2008))

- A broader participation of financial operators increases the interdependence among the futures prices of different asset classes (e.g. Büyükşahin et al. (2009a), Büyükşahin et al. (2009b), Büyükşahin and Harris (2011), Tang and Xiong (2012), Büyükşahin and Robe (2014) and Silvennoinen and Thorp (2013))

On the effects of speculation on oil prices

- A large inflow of financial investors in oil futures markets is associated with a surge of spot and futures crude oil prices (e.g. Fattouh et al. (2013), Alquist and Kilian (2010)).

    - Many researchers employ SVAR models to identify the impact of speculation on crude oil price changes (e.g. Kilian and Murphy (2012)).

    - More sophisticated econometric techniques (e.g. Multivariate GARCH models) are employed by, e.g., Du et al. (2011), Cifarelli (2013), Manera et al. (2013).

Data

We focus on the US crude oil market:

Weekly data from Feb-2000 to Nov-2014 (771 weekly observations):

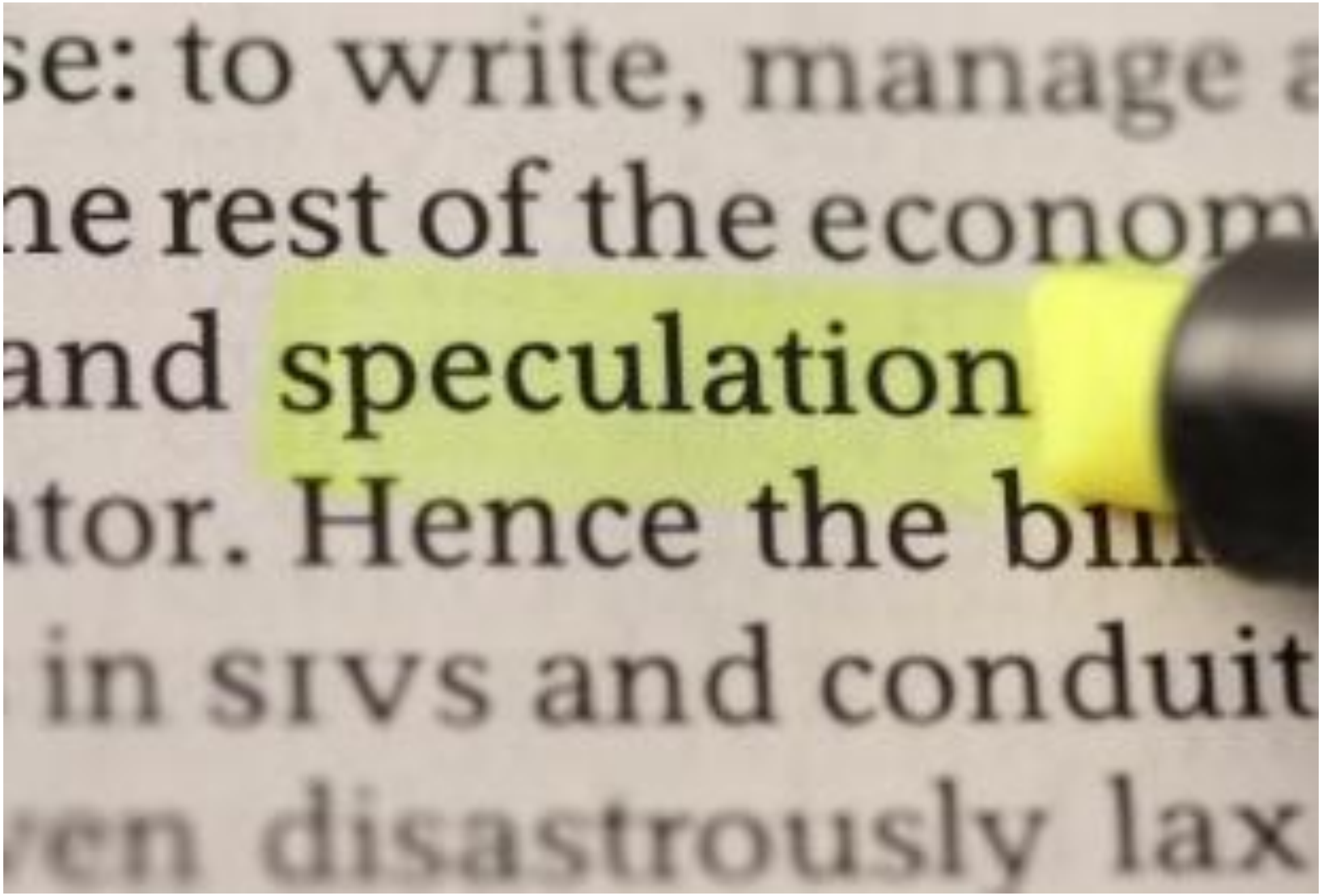
- Fundamental data:
  EIA US supply-demand of crude oil and petroleum products.

- Market data:
  - NYMEX WTI front month close price (WTI_1_t).
  - NYMEX timespreads (1^{st}-4^{th} month).

  - CFTC commitment of traders (COT) report data:
    - Commercial positions (“hedgers”).
    - Non commercial plus non reporting (“speculators”).

We also introduce a measure for the **equilibrium price of crude oil** based on Saudi Arabia’s breakeven price.

What is speculation?



Measures of speculation

Working's T-Index:

$$T-Index_t = \begin{cases} 1+\frac{SS_t}{CS_t+CL_t} & \text{if } CS_t \geq CL_t \\ 1+\frac{SL_t}{CS_t+CL_t} & \text{if } CS_t < CL_t \end{cases}$$

Percent of total open interest held by each CFTC trader classification:

$$\text{Reporting Commercial percent of } TOI_t = \frac{CL_t + CS_t}{2 \cdot TOI_t}$$

Percent net long (PNL) position:

$$PNL_t^{Com} = \frac{CL_t - CS_t}{CL_t + CS_t}$$

$$PNL_t^{Spec} = \frac{SL_t - SS_t}{SL_t + SS_t}$$

where

$$SL_t = NCL_t + NRL_t + SP_t$$

$$SS_t = NCS_t + NRS_t + SP_t$$

Lakonishok, Shleifer and Vishny (1992)'s H-Index:

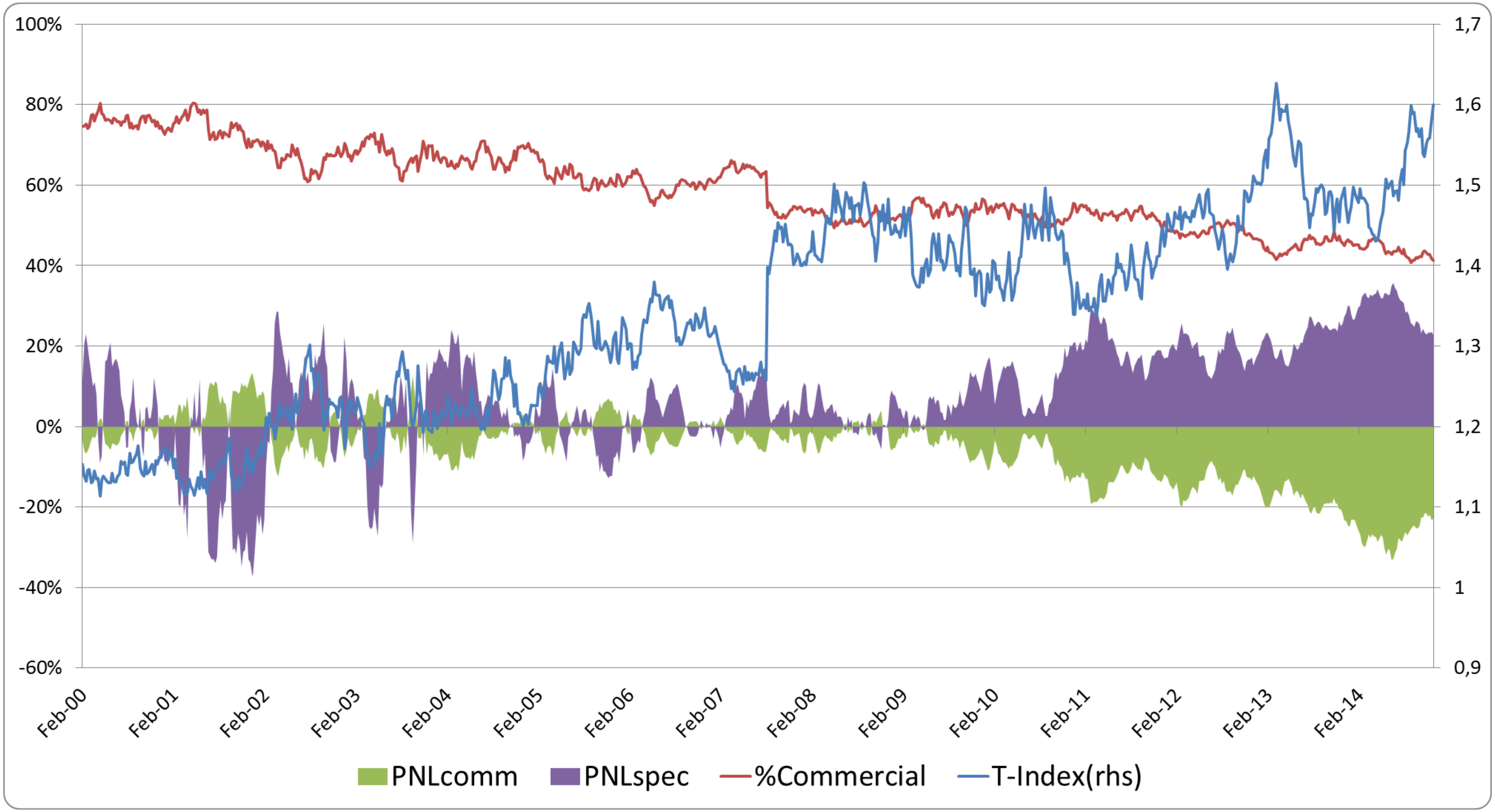
$$H_t = \left| \frac{B_t}{B_t + S_t - p_t} \right| - AF_t$$

Measured related to volume and open interest data:

$$\text{(a) } R1_t = \frac{V_t}{TOI_t}, \text{ (b) } R2_t = \frac{V_t}{|\Delta TOI_t|} \text{ (c) } R3_t = \frac{\Delta TOI_t}{V_t}$$

Legenda: NCS_t (resp. NCL_t): non-commercial short (resp. long); NRS_t (resp. NRL_t): non-reporting short (resp. long); SS_t (resp. SL_t): speculation short (resp. long); CS_t (resp. CL_t): reporting commercial net short (resp. long) positions; SP_t: amount of spreading positions; p_t: expected proportion of money managers buying relative to the number of active; B_t (resp. S_t): net buyers (resp. sellers); AF_t: adjustment factor; V_t: total volume of contracts traded and TOI_t: market's total open interest.

Measures of speculation



Speculative OI has overtaken commercial OI while '*speculators*' have apparently become structurally long.

Our measure of speculation

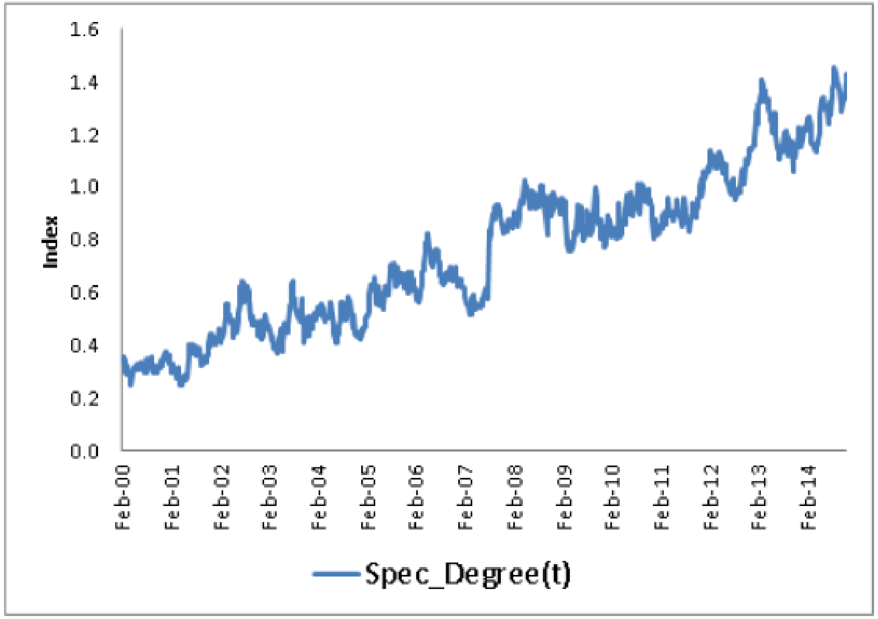
- **<u>Speculation indexes</u>** show that the importance of speculation in the WTI derivatives market has strongly increased:

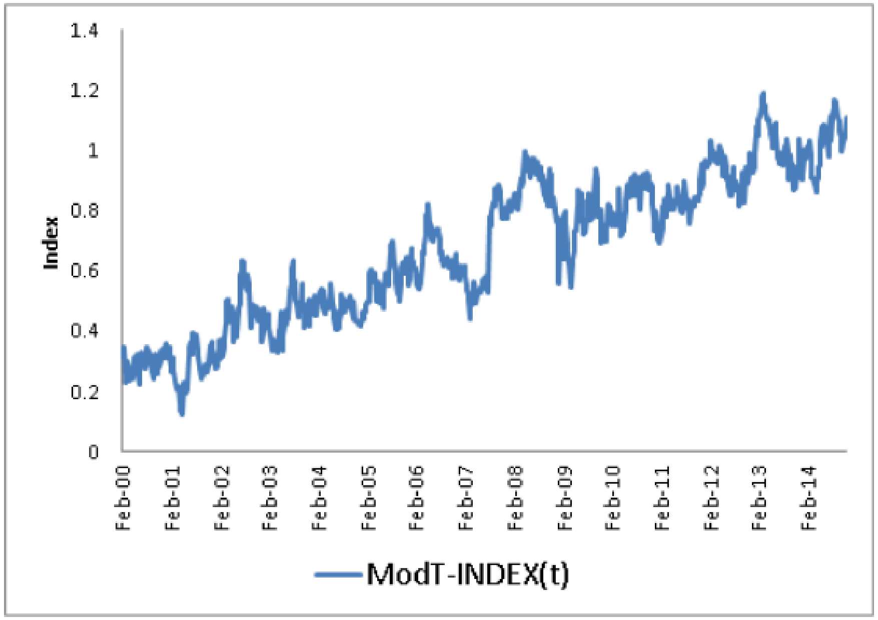
$$\text{Spec\_Degree}_t = \frac{OI^{spec}}{OI^{comm}}$$

(ratio between speculation and commercial open interest, $spec/comm$)

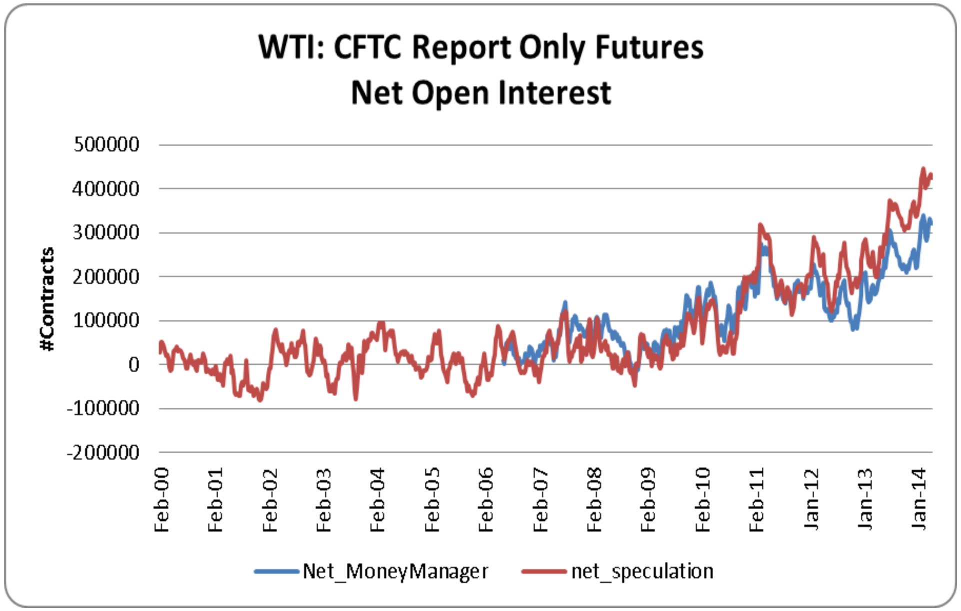
$$\text{Modified}_{\text{T-Index}}$$

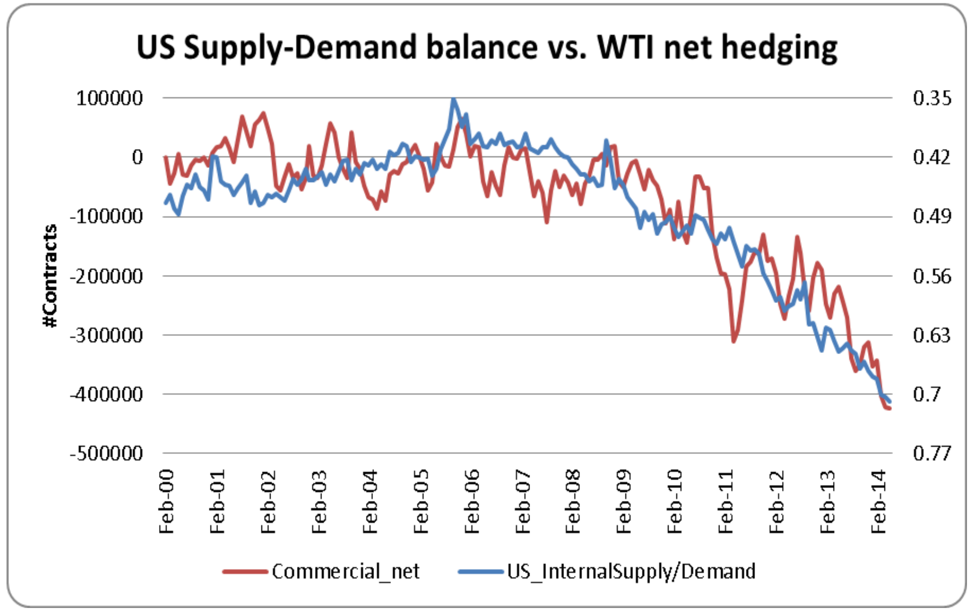
(a version of the famous index modified taking into account our estimation of hedging needs).





Graphical analysis (1/4)





- The rise of non commercial net long positions can be almost entirely attributed to the increasing presence of **money managers** on the future crude oil market.

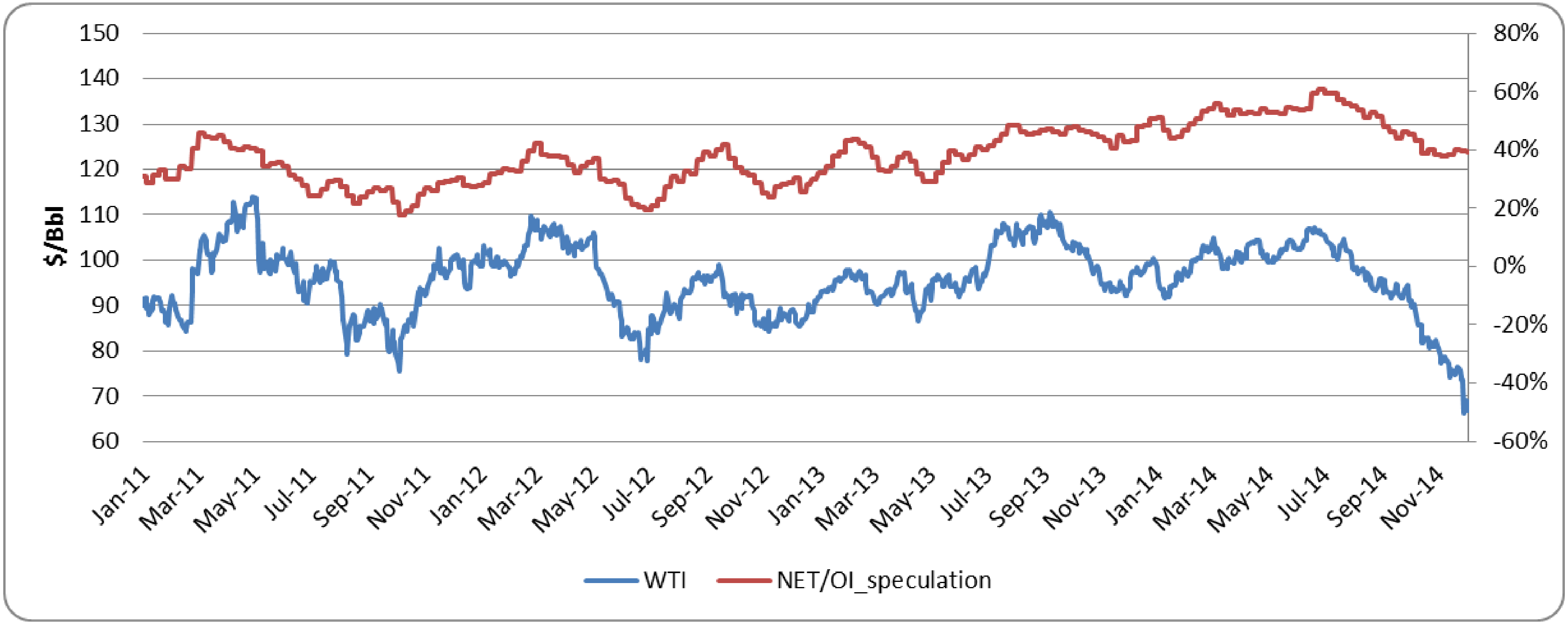
- The expansion in **US crude production** is changing the hedging needs of operators.

EDISON

Market Analysis and Forecasting

Graphical analysis (2/4)

- The huge increase in net long positions seems **not** to be associated with a corresponding increase in the price of oil.

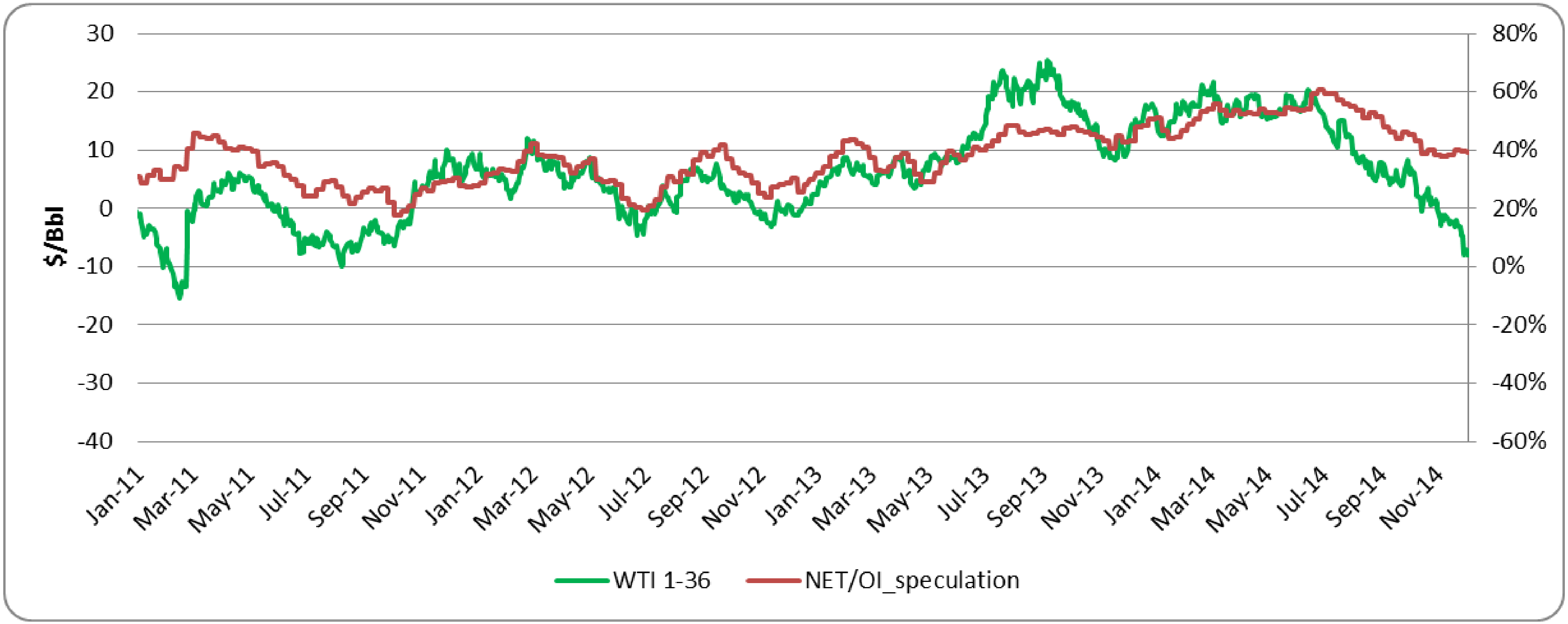


EDISON

Market Analysis and Forecasting

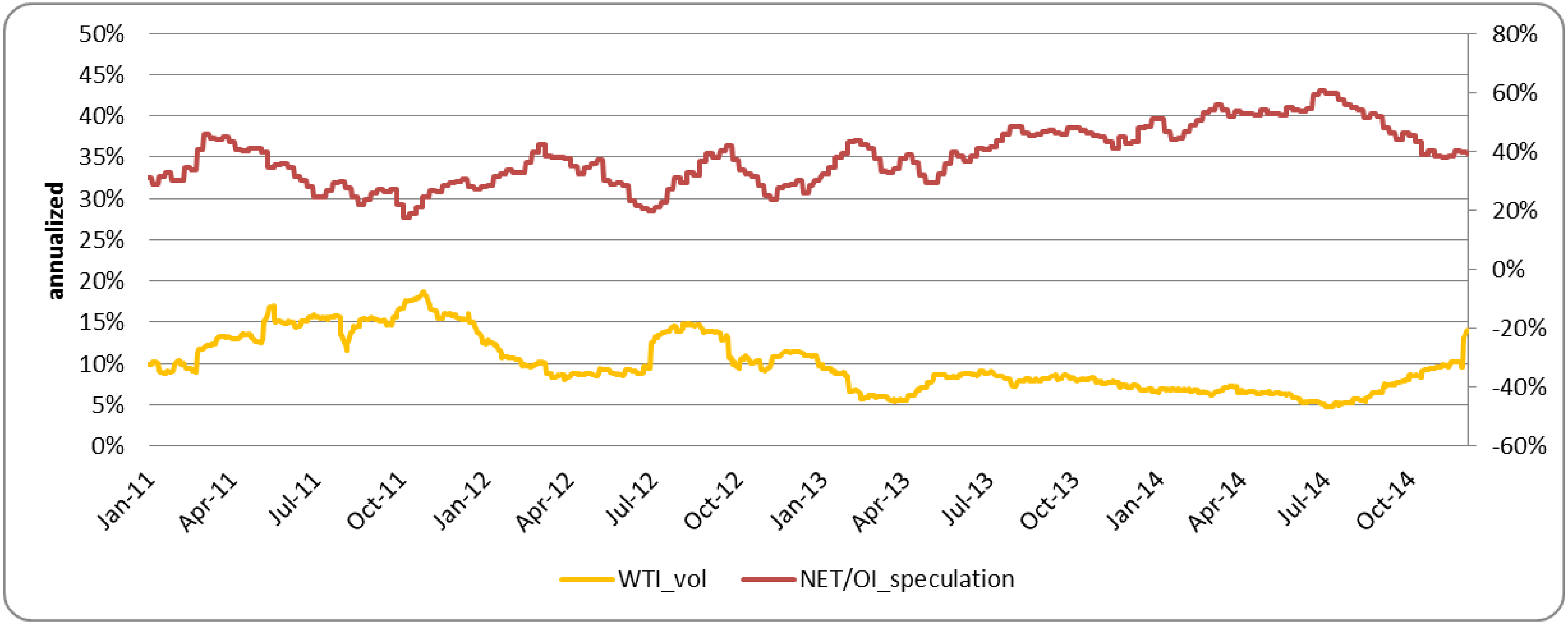
Graphical analysis (3/4)

- Backwardation has, on the contrary, considerably risen (especially on the back-end of the curve), with the exception of the contango that characterized the last part of the sample.



Graphical analysis (4/4)

- Despite high geopolitical risk, volatility has sensibly decreased.



For all these reasons we could attribute to "*speculators*" a NEW and increasingly important role as they effectively contribute to stabilize oil prices and reduce volatility.

EDISON

Market Analysis and Forecasting

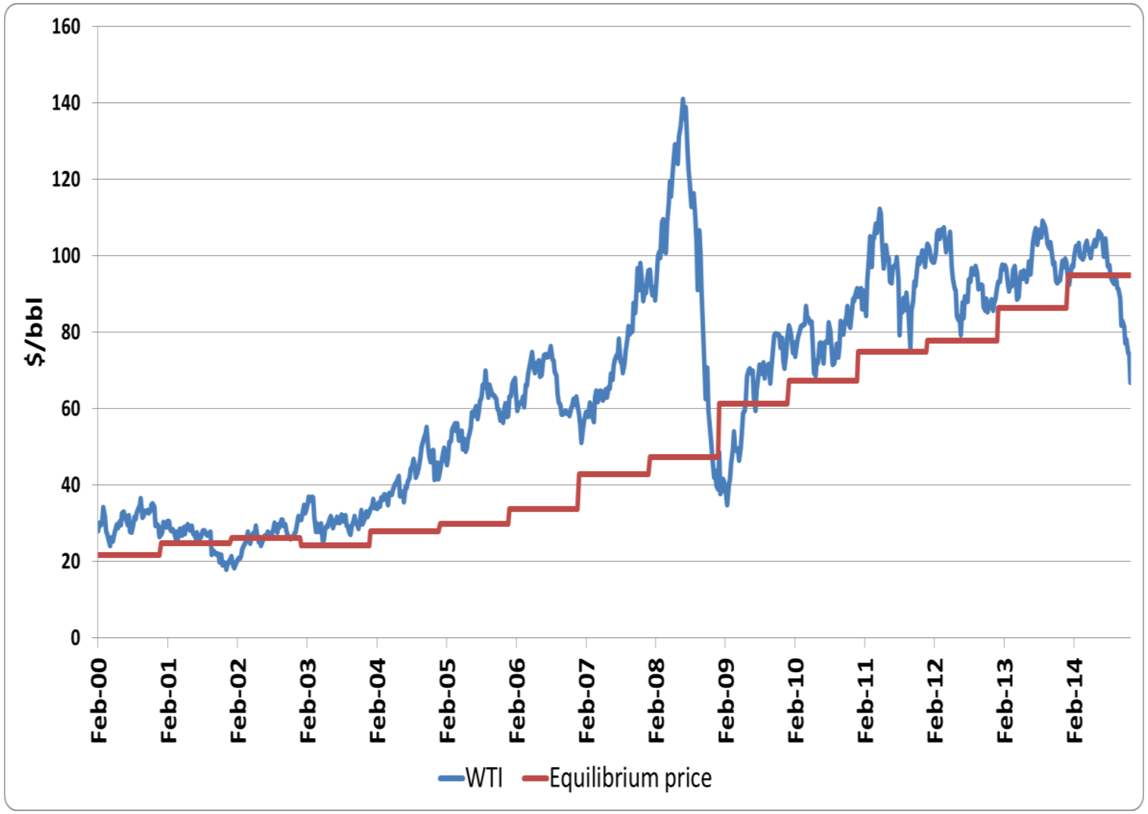
The equilibrium price (1/2)

- We believe the price should fluctuate around a **long run equilibrium.**
- Saudi Arabia plays a key role in the supply arena. Since the foundation of OPEC the Kingdom has tried to adjust production in order to protect oil revenues.
- So we argue that the Saudi Breakeven price can be considered as a sort of equilibrium

$$Gov\_Rev_t = f(x_t) = f(Prod_t \cdot P_t)$$

$$\widehat{P}_t = \widehat{f}^{-1}(Gov\_Exp_t)$$

Year	Equilibrium Price ($/bbl)
2011	74.91
2012	77.89
2013	86.41
2014	94.91



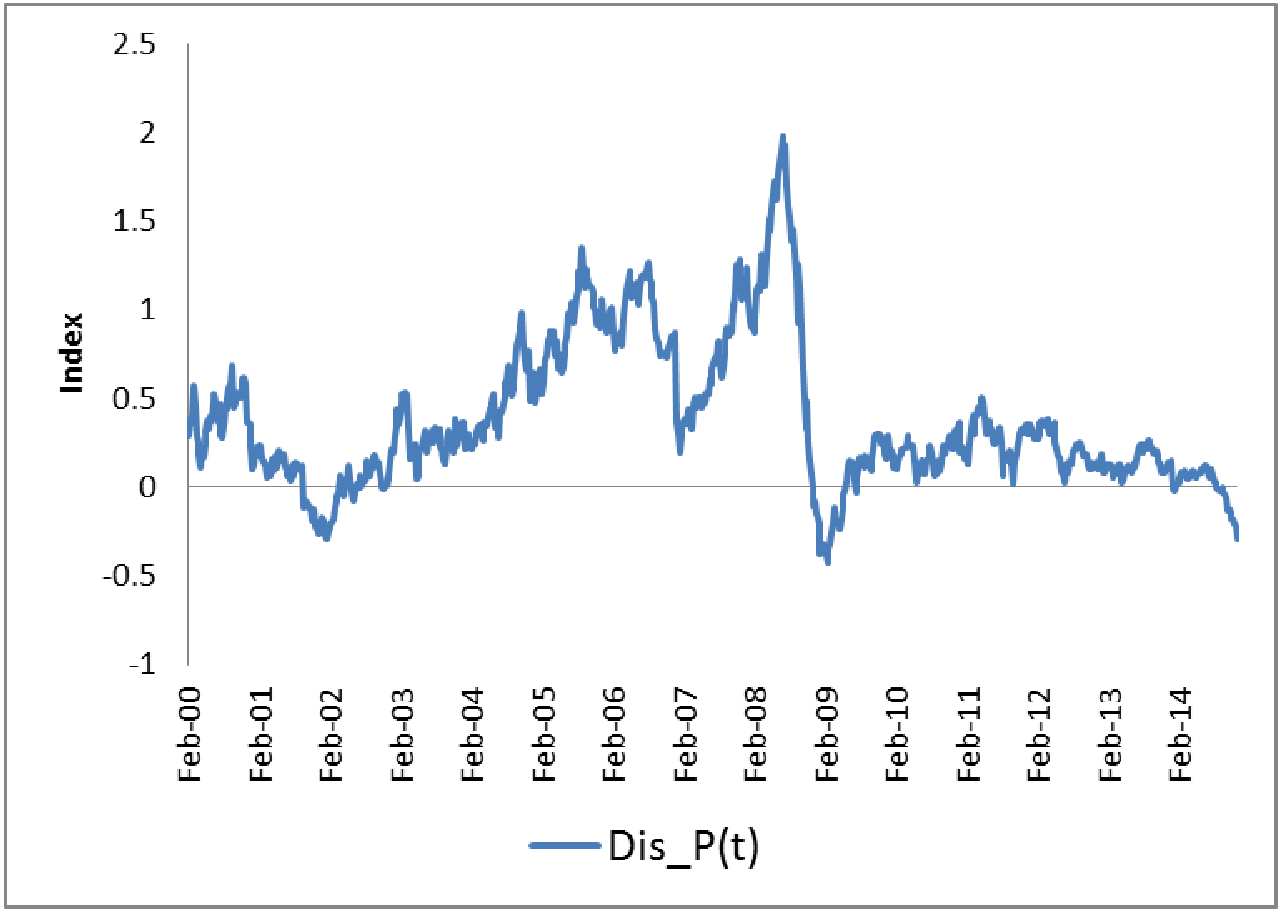
Under **equilibrium:**

- Crude oil supply-demand balance is stable.

The equilibrium price (2/2)

- The **disequilibrium price** is computed as:

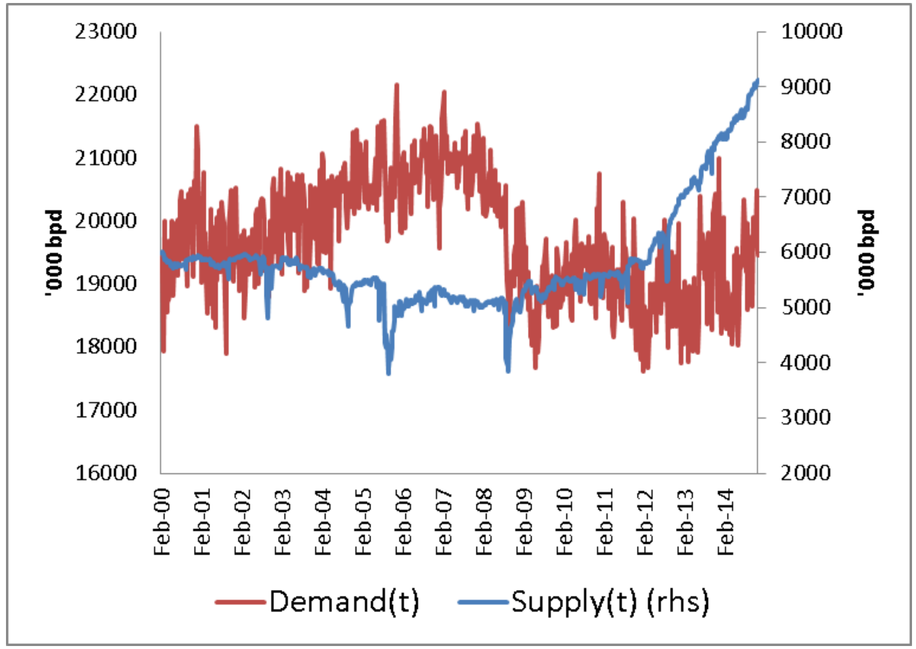
$$Dis\_P_t = \frac{WTI\_1_t - \widehat{P}_t}{\widehat{P}_t}$$



Our variables

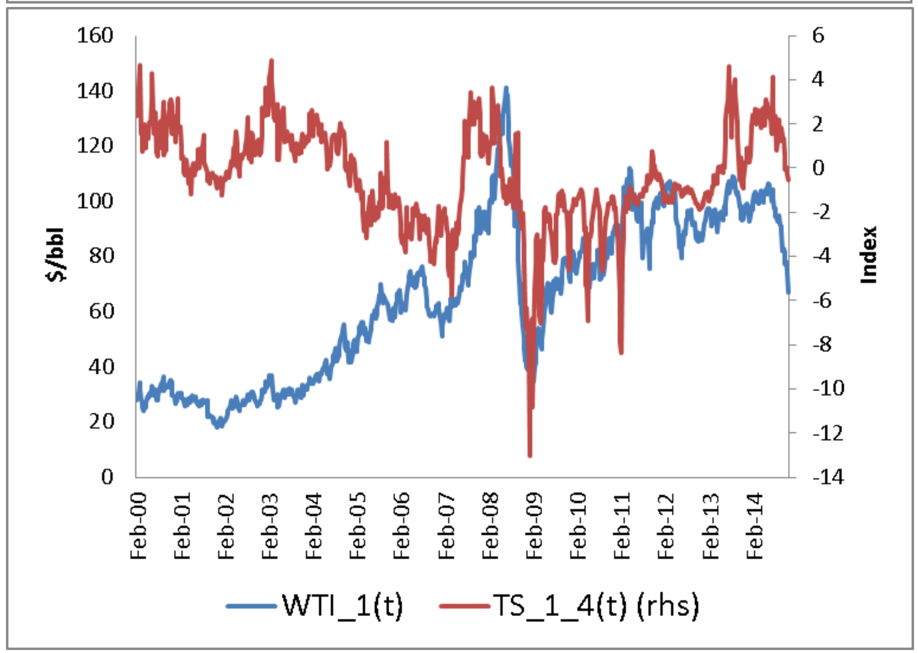
- US crude oil supply-demand balance:

$$Bal_t^{US} = \frac{Demand_t - Supply_t}{Demand_t + Supply_t}$$



- WTI 1st-4th month timespread:

$$TS\_1\_4_t = \frac{WTI\_1_t - WTI\_4_t}{WTI\_1_t}$$

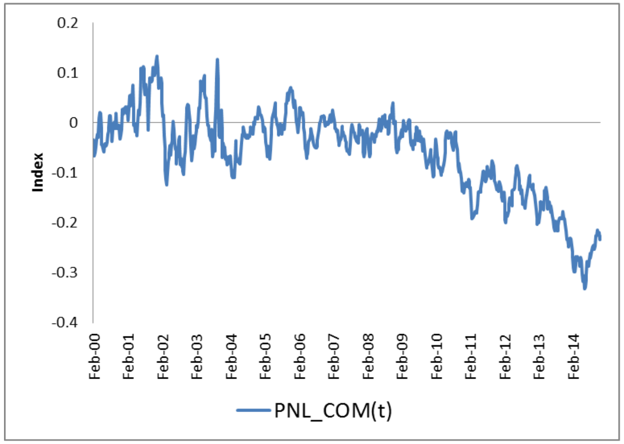


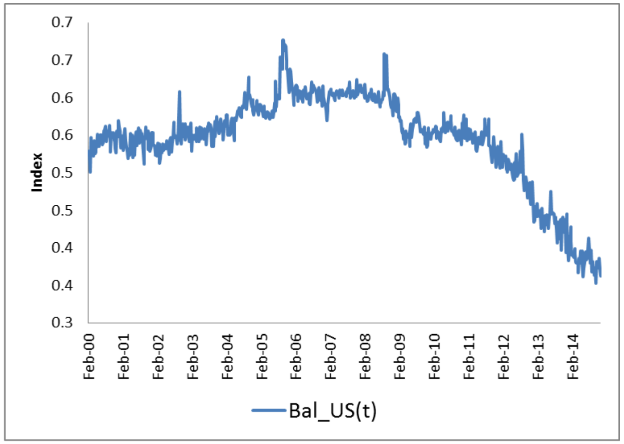
EDISON

Market Analysis and Forecasting

Pesaran et al. (2001)'s procedure (1/4)

- **1st step.** *Statistical properties of the data.* The order of integration of series is examined through Augmented Dickey Fuller (ADF) unit root tests.





- Mixture of I(1), I(0) series.
- No I(2) series.

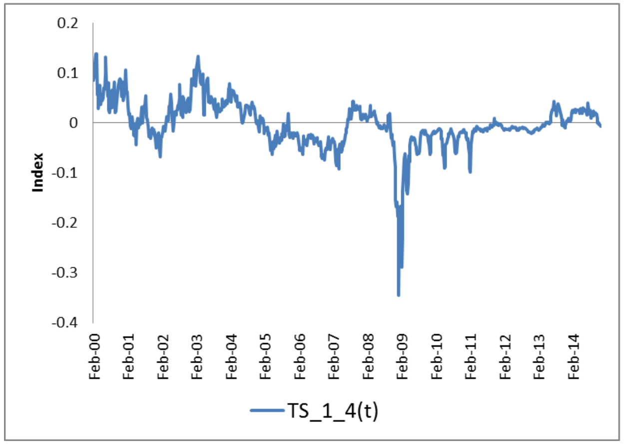


Table 2: **ADF tests for unit root**

Regressor	**Level**		**First Difference**		**Result**
	***t*-statistics**	**Prob.**	***t*-statistics**	**Prob.**	
PNL_t^{Com}	-2.667	0.080*	-23.547	0.000***	$I(1)$
Bal_t^{US}	-0.049	0.953	-23.385	0.000***	$I(1)$
$TS\_1\_4_t$	-4.067	0.001***	-20.334	0.000***	$I(0)$
Dis_P_t	-2.125	0.235	-28.176	0.000***	$I(1)$

Notes: This table reports the unit root tests using the ADF tests for the set of variables considered by the ARDL analysis. ***, ** and * denote rejection of the null hypothesis of unit root at the 1%, 5% and 10% significance level, respectively.

Pesaran et al. (2001)'s procedure (2/4)

- **2nd step.** *Estimation of a Conditional Error Correction Model (ECM).*

$$\Delta Y_t = \beta_0 + \theta_0 Y_{t-1} + \theta_1 X_{1,t-1} + \theta_2 X_{2,t-1} + \theta_3 X_{3,t-1} + \sum_{l=1}^{k_1} \beta_{1,l} \Delta Y_{1,t-l} + \\ + \sum_{l=1}^{k_1} \beta_{2,l} \Delta X_{2,t-l} + \dots + \sum_{l=1}^{k_n} \beta_{n,l} \Delta X_{n,t-l} + \epsilon_t \quad (1)$$

where

- $Y_t = PNL_t^{COMM}$
- $X_{1,t} = Bal_t^{US}$
- $X_{2,t} = TS\_1\_4_t$
- $X_{3,t} = Dis\_Pt$ ⇐ Not statistically significant

EDISON

Market Analysis and Forecasting

Pesaran et al. (2001)'s procedure (3/4)

Conditional ECM:

- The appropriate number of lags is selected by using the Schwartz information criterion. ✓
- Check significance of variables in levels. ✓
- Check for serial dependence of regression errors (LM test). ✓

Table 3: ARDL Model with Dependent variable ΔPNL_t^{Com}

Variable	full model		reduced model	
	value	prob	value	prob
c	-0.043	0.000***	-0.035	0.000***
PNL_{t-1}^{Com}	-0.052	0.000***	-0.048	0.000***
Bal_{t-1}^{US}	0.072	0.000***	0.059	0.000***
$TS\_1\_4_{t-1}$	0.010	0.003***	0.028	0.023**
Dis_P_{t-1}	-0.001	0.457		
ΔPNL_{t-1}^{Com}	0.130	0.001***	0.181	0.000***
ΔBal_{t-1}^{US}	-0.002	0.956	0.002	0.966
$\Delta TS\_1\_4_{t-1}$	-0.017	0.321	-0.046	0.160
ΔDis_P_{t-1}	-0.028	0.006***		
F-statistic	8.437	0.000***	8.191	0.000***
Akaike info criterion	-5.514		-5.496	
Schwarz criterion	-5.459		-5.454	
Durbin-Watson stat	2.016		1.989	
Breusch-Godfrey Serial Correlation LM Test	2.238	0.107	0.530	0.589

Notes: This table shows the estimated coefficients and diagnostic properties of residuals of the AutoRegressive Distributed Lag Model (1). ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

Pesaran et al. (2001)'s procedure (4/4)

Table 4: **Pesaran et al. [29]'s bounds tests**

Wald Test			Critical bounds ($k = 2$)		
Null hypothesis $H_0 : \theta_0 = \theta_1 = \theta_2 = 0$			**Unrestricted intercept and no trend**		
Test Statistic	*df*	Value	Prob.	$I(0)$	$I(1)$
F-Statistic	(3, 764)	8.611***	0.01	5.15	6.36
Wald Test			**Critical bounds ($k = 2$)**		
$H_0 : \theta_0 = 0$, $H_1 : \theta_0 < 0$			**Unrestricted intercept and no trend**		
Test-statistic	*df*	Value	Prob.	$I(0)$	$I(1)$
t-Statistic	(764)	-4.824***	0.01	-3.43	-4.10

Notes: This table shows the results obtained by following the ARDL approach presented in Pesaran et al. [29]. The estimated coefficients refer upon the conditional ECM given by eq. (2). ***, ** and * denote rejection of the null hypothesis at the 1%, 5% and 10% significance level, respectively.

- **3rd step.** *Pesaran et al. (2001)'s "bounds tests".* Given equation (1), the hypothesis of the absence of a long-run equilibrium relationship among the variables is tested against the presence of a cointegration vector.

Null hypothesis: $H_0: \theta_0 = \theta_1 = \theta_2 = 0$

Alternative hypothesis: H_0 is not true

$\theta_0 \longrightarrow PNL^{Com}_{t-1}$

$\theta_1 \longrightarrow Bal^{US}_{t-1}$

$\theta_2 \longrightarrow TS\_1\_4_{t-1}$

Rejection of the null hypothesis implies the presence of a long-run cointegrating relationship.

Hedging needs model

Estimation of the full structural VECM:

Since we validated the existence of a long run equilibrium we are able to estimate the system of equations:

$$\Delta X_{i,t} = C_{i01}\Delta X_{1,t-1} + C_{i02}\Delta X_{2,t-1} + C_{i03}\Delta X_{3,t-1} + \quad (2)$$
$$+C_{i06}\left(X_{1,t-1} - LC_1 - LC_2 X_{2,t-1} - LC_3 X_{3,t-1}\right) + \varepsilon_{it}$$

where

- $i = 1, 2, 3$
- $X_{1,t} = PNL_t{}^{Com}$
- $X_{2,t} = Bal_t^{US}$
- $X_{3,t} = TS\_1\_4_t$
- LC_1, LC_2 and LC_3 denote the long-run coefficients

The system incorporates the long-run equilibrium vector as well as the dynamic relationships among the variables of interest (Vector Error Correction model).

Hedging needs – Full structural VECM

$$Bal_t^{US} = \frac{Demand_t - Supply_t}{Demand_t + Supply_t}$$

$$TS\_1\_4_t = \frac{WTI_{1_t} - WTI\_4_t}{WTI\_1_t}$$

$$PNL_t^{Com} = \frac{CL_t - CS_t}{CL_t + CS_t}$$

In the short-run Bal_t^{US} has a statistically significant impact on $TS\_1\_4_t$

Table 5: VECM Estimates: Model for Hedging Needs

Coefficient	Value	Std. Error	t–Statistic	Prob.
C_{101}	0.176	0.036	4.946	0.000***
C_{102}	0.003	0.043	0.081	0.936
C_{103}	-0.052	0.032	-1.590	0.112
C_{106}	-0.039	0.009	-4.202	0.000***
LC_1	-0.808	0.108	-7.509	0.000***
LC_2	1.381	0.197	7.006	0.000***
LC_3	1.043	0.308	3.385	0.001***
C_{201}	0.003	0.029	0.110	0.913
C_{202}	-0.313	0.034	-9.106	0.000***
C_{203}	0.001	0.026	0.045	0.964
C_{206}	0.014	0.007	2.074	0.038**
C_{301}	0.050	0.039	1.276	0.202
C_{302}	0.084	0.047	1.784	0.075*
C_{303}	-0.163	0.036	-4.560	0.000***
C_{306}	0.030	0.010	3.139	0.002***

Included observations: 771

Total system (balanced) observations 2313

Notes: This table shows the results of the VECM estimates (system of equations (2)). ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

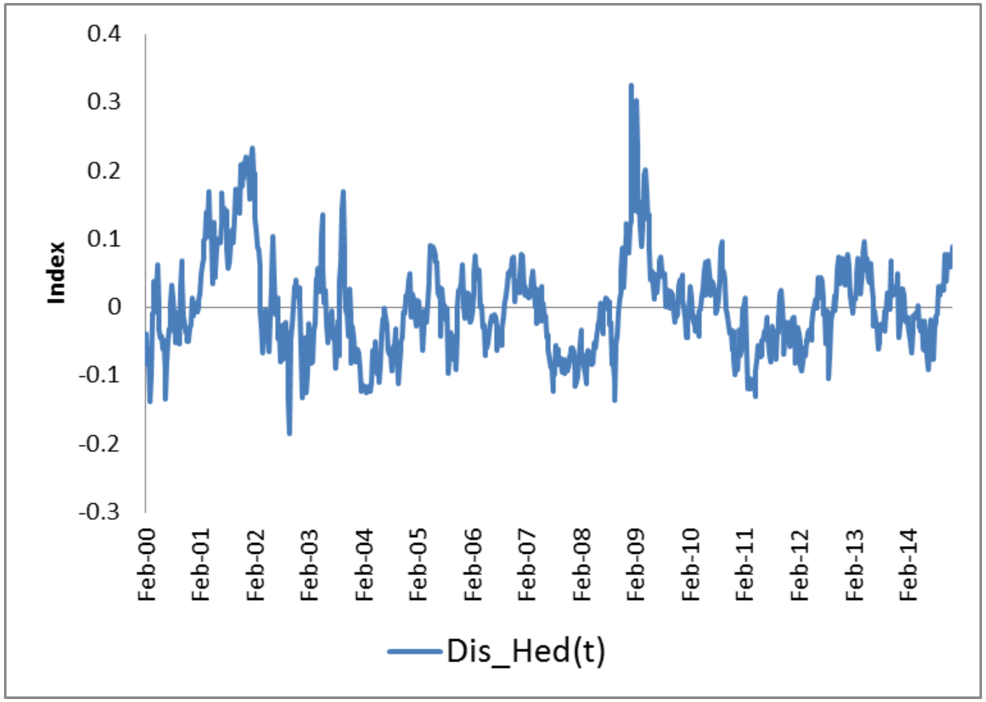
Long-run coefficients:

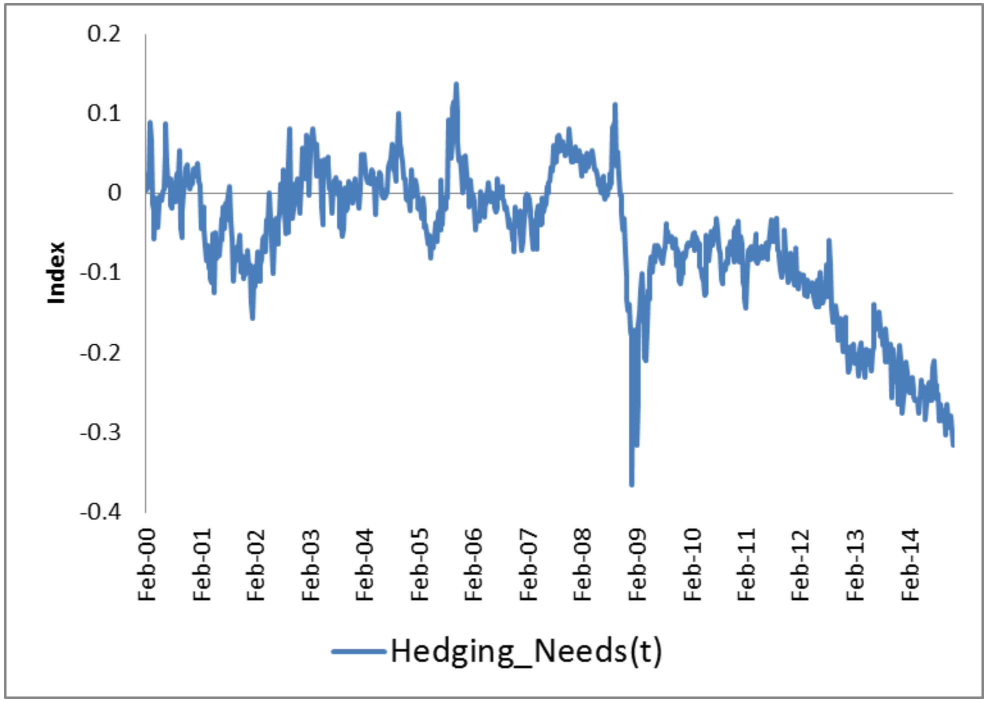
a) As Bal_t^{US} decreases, the hedging needs decrease as well ($LC_2 > 0$)

b) There is also a direct relationship between $TS\_1\_4_t$ and operators' hedging needs ($LC_3 > 0$)

The statistical relevance of coefficients $C_{106}, C_{206}, C_{306}$ denote an adjustment mechanism towards equilibrium of the three variables

Hedging needs and disequilibrium hedging





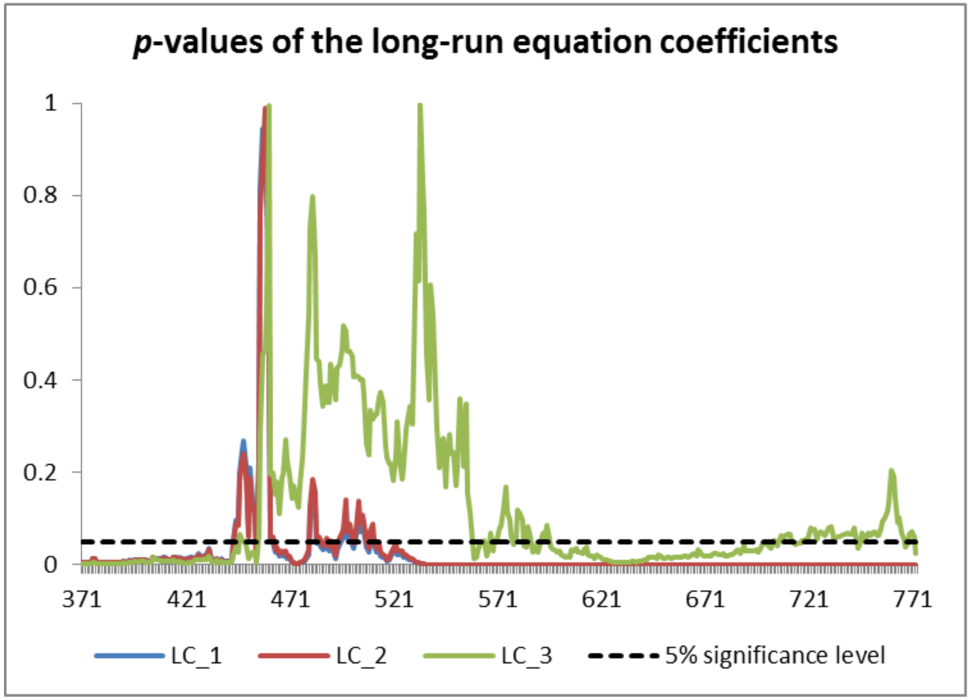
$$Hedging\_Needs_t = \widehat{LC}_1 + \widehat{LC}_2 \cdot Bal_t^{US} + \widehat{LC}_3 \cdot TS\_1\_4_t \qquad (3)$$

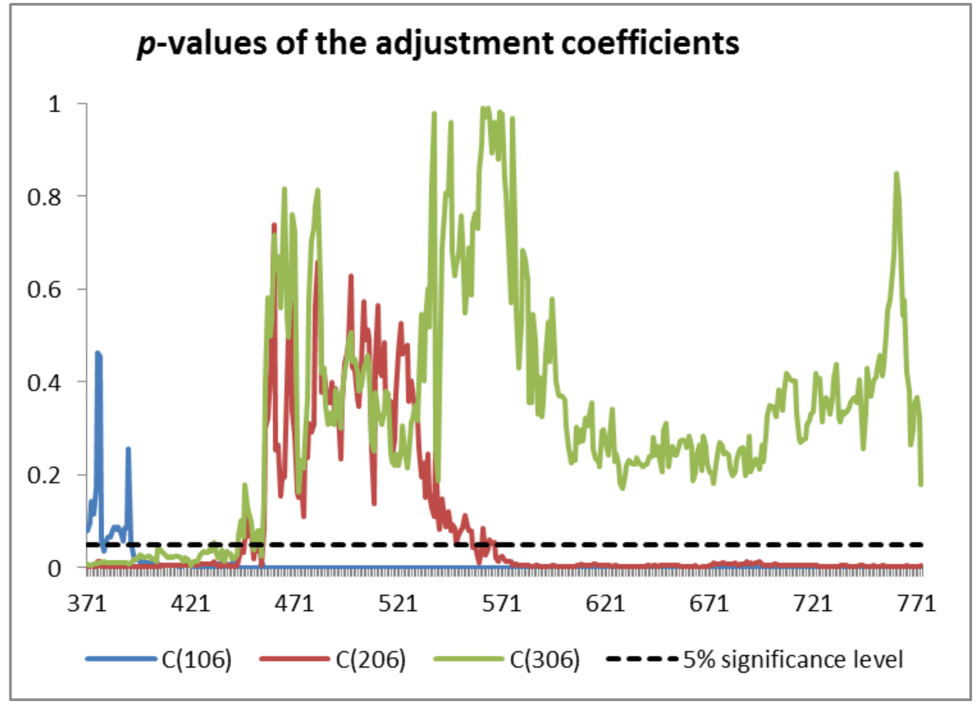
The corresponding "disequilibrium hedging" (Dis_Hed_t) is given by:

$$Dis\_Hed_t = PNL_t^{Com} - Hedging\_Needs_t$$

$$Mod\,T - INDEX_t = \frac{(Speculators\_OI_t - |Hedging\_Needs_t \cdot Commercial\_OI_t|)}{Commercial\_OI_t}$$

Robustness of the model





A sequence of rolling windows of fixed length of 371 obs (around 7 years of data) is used to investigate whether coefficients of the long run equilibrium equation (3) (parameters LC_1, LC_2 and LC_3) do **not** vary across samples.

EDISON

Market Analysis and Forecasting

Price model (1/5)

$$\Delta Dis\_P_t = C_1 + \omega_0 Dis\_P_{t-1} + C_2 \Delta Dis\_P_{t-1} + C_3 \Delta PNL_{t-1}^{Com} + \qquad (4)$$
$$+ C_4 \Delta Bal_{t-1}^{US} + C_5 \Delta TS\_1\_4_{t-1} + C_6 \Delta Dis\_Hed_{t-1} + \omega_1 \cdot S_{t-1} + \eta_t$$

here S_t denotes a variable obtained by multiplying Dis_P_t by a measure of speculative activity on the market; that is, $S_t = Dis\_P_t \cdot Spec\_var_t$ where $Spec\_var_t$ is either $Spec\_Degree_t$ or $Mod\,T - INDEX_t$.

Price model (2/5)

Table 6: Oil price model ($Spec\_var_t = Spec\_Degree_t$): Full Model

Variable	Coefficient	Std. Error	$t-$Statistic	Prob.
c	0.049	0.006	7.894	0.000***
ΔDis_P_{t-1}	-0.001	0.053	-0.026	0.979
ΔPNL^{Com}_{t-1}	0.262	0.204	1.287	0.199
ΔBal^{US}_{t-1}	-0.072	0.226	-0.318	0.751
$\Delta TS\_1\_4_{t-1}$	0.157	0.243	0.645	0.519
ΔDis_Hed_{t-1}	-0.064	0.057	-1.119	0.264
S_{t-1}	-0.146	0.040	-3.626	0.000***
Dis_P_{t-1}	-0.073	0.026	-2.859	0.004***
R-squared	0.146	Mean dependent var		0.001
Adjusted R-squared	0.133	S.D. dependent var		0.068
S.E. of regression	0.063	Akaike info criterion		-2.668
Sum squared resid	1.808	Schwarz criterion		-2.597
Log likelihood	623.065	Hannan-Quinn criter.		-2.640
F-statistic	11.100	Durbin-Watson stat		1.833
Prob(F-statistic)	0.000***			

Dependent Variable: ΔDis_P_t
Method: Least Squares
Sample: $q(0.2) < Dis\_P_t < q(0.8)$
Included observations: 461

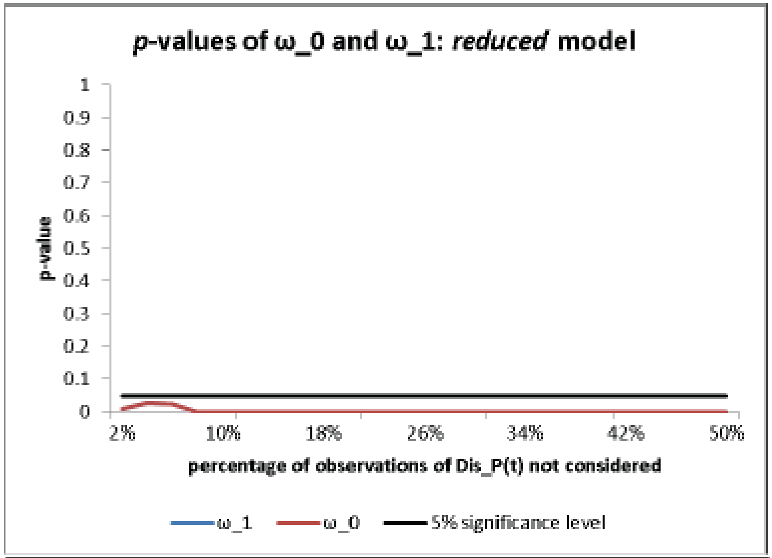
Notes: This table shows the results obtained by estimating equation (4). ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

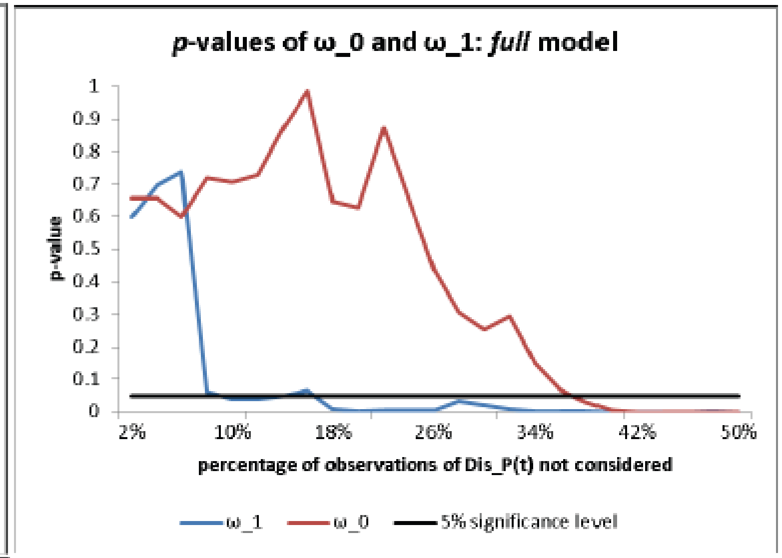
Price model (3/5)

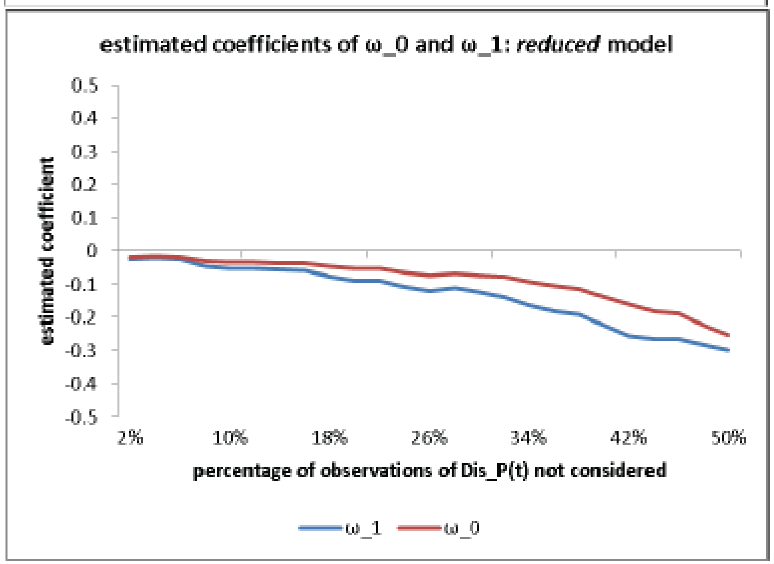
Table 7: Oil price model ($Spec\_var_t = ModT - INDEX_t$): Full Model

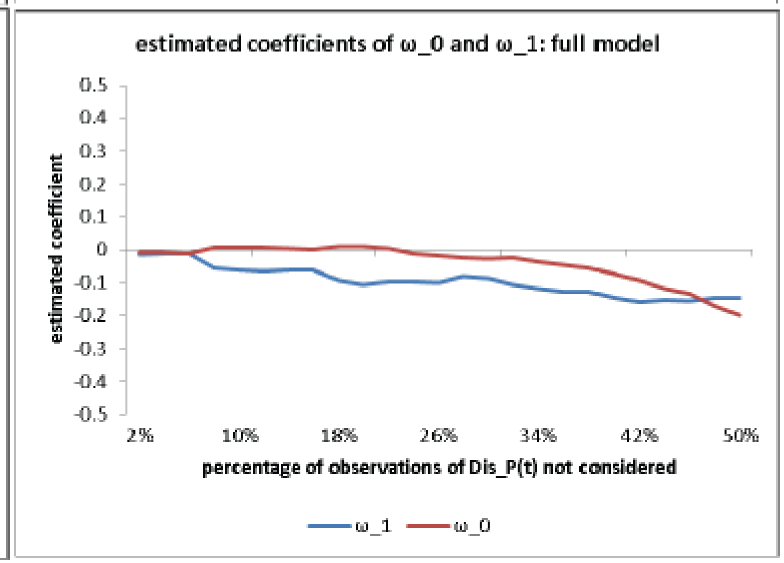
Variable	Coefficient	Std. Error	$t-$Statistic	Prob.
c	0.047	0.006	7.655	0.000***
ΔDis_P_{t-1}	-0.002	0.054	-0.044	0.965
ΔPNL^{Com}_{t-1}	0.258	0.205	1.260	0.208
ΔBal^{US}_{t-1}	-0.058	0.226	-0.256	0.798
$\Delta TS\_1\_4_{t-1}$	0.157	0.244	0.645	0.520
ΔDis_Hed_{t-1}	-0.058	0.058	-1.015	0.311
S_{t-1}	-0.138	0.045	-3.095	0.002***
Dis_P_{t-1}	-0.078	0.027	-2.930	0.004***
R-squared	0.140	Mean dependent var		0.001
Adjusted R-squared	0.127	S.D. dependent var		0.068
S.E. of regression	0.063	Akaike info criterion		-2.661
Sum squared resid	1.822	Schwarz criterion		-2.589
Log likelihood	621.293	Hannan-Quinn criter.		-2.632
F-statistic	10.519	Durbin-Watson stat		1.832
Prob(F-statistic)	0.000***			
Dependent Variable: ΔDis_P_t				
Method: Least Squares				
Sample: $q(0.2) < Dis\_P_t < q(0.8)$				
Included observations: 543				
Notes: See notes to Table 6.				

Price model (4/5) – Spec_Degree

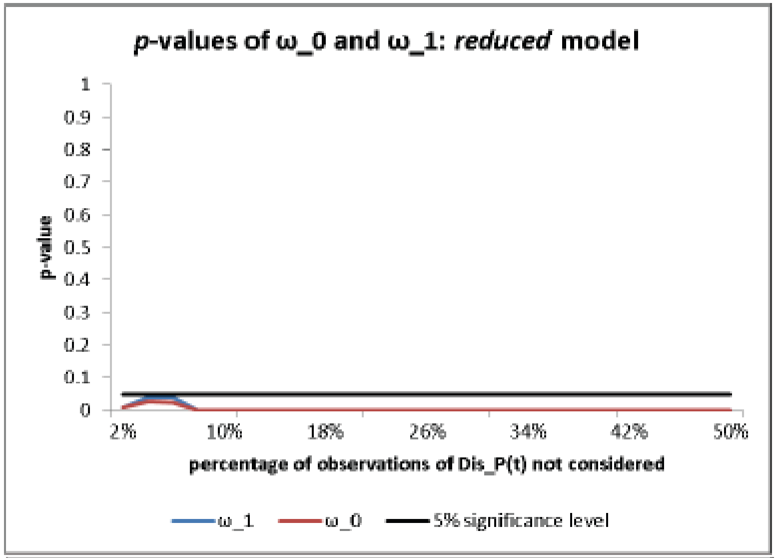


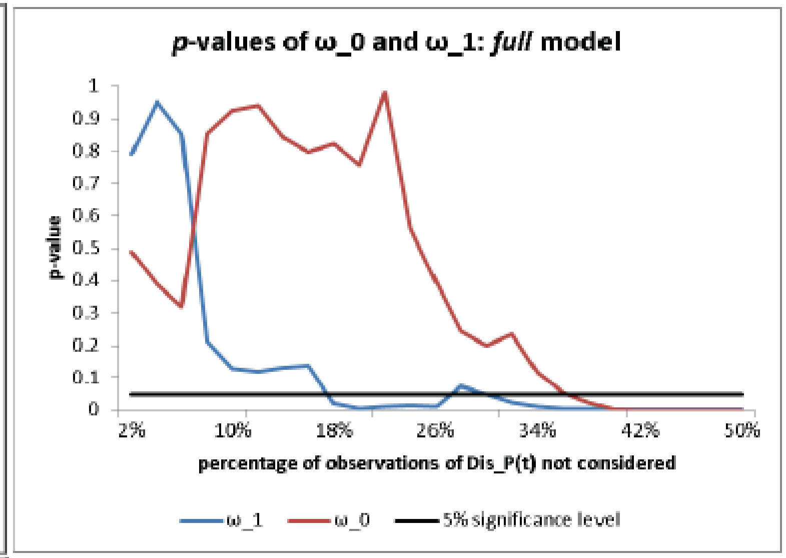


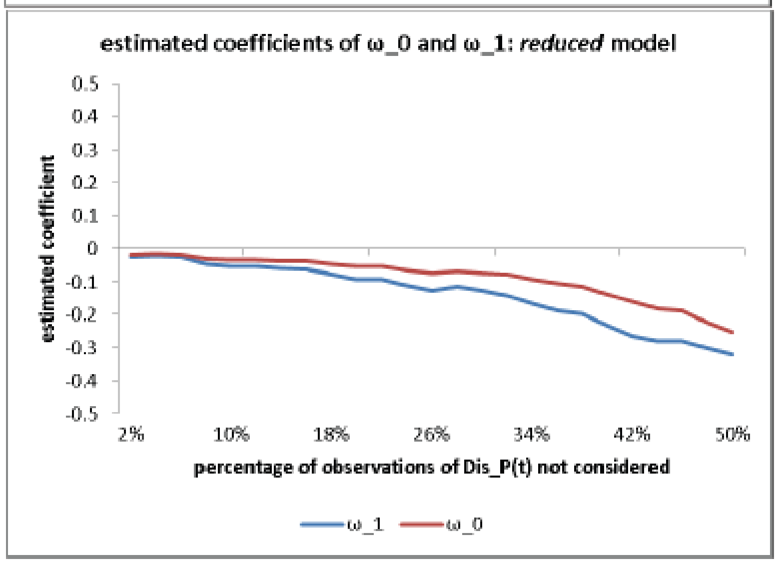


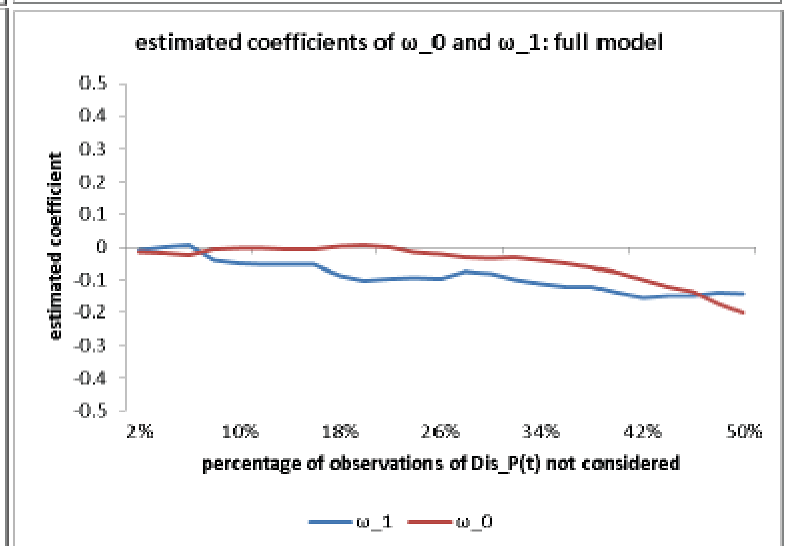


Price model (5/5) – Mod_T_Index







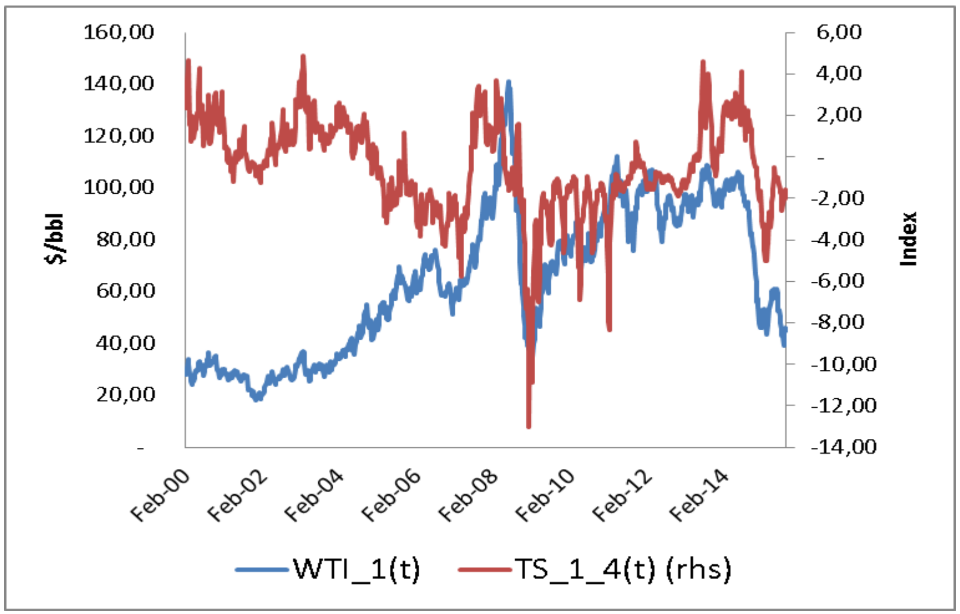


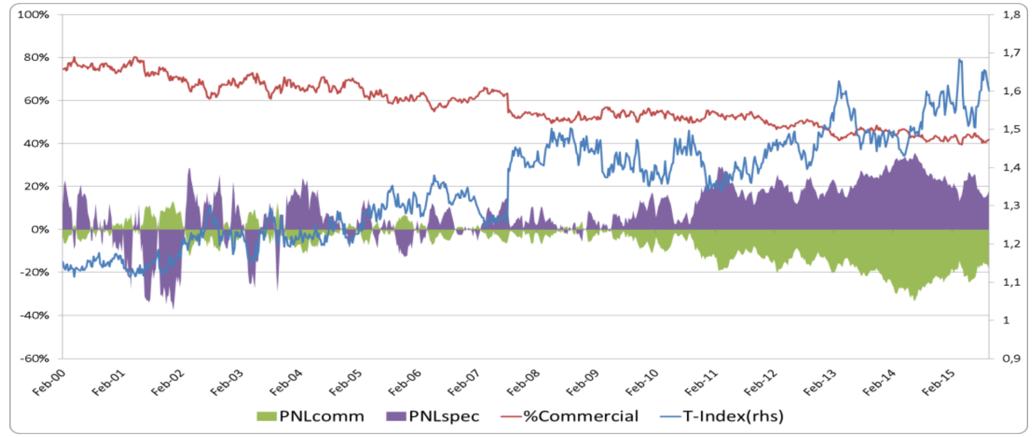
Conclusions

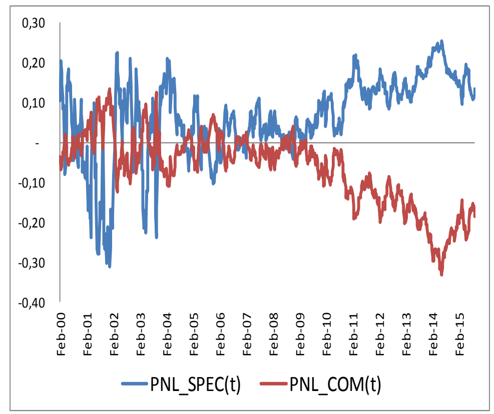
- We examine the relationship between positions held by commercial operators, crude oil prices and US supply-demand balance.
- We provide new insights on the economics of market participants and on the functioning of US crude oil market.
- In particular, we analyze the effects of speculation on the adjustment mechanisms of crude oil prices towards equilibrium.
- We find empirical evidence of **speculators stabilizing crude oil prices**.
- Non-commercial operators play a fundamental role in international crude oil markets.
- The model is robust and takes into account a vast sample (with shocks) and includes very recent observations.

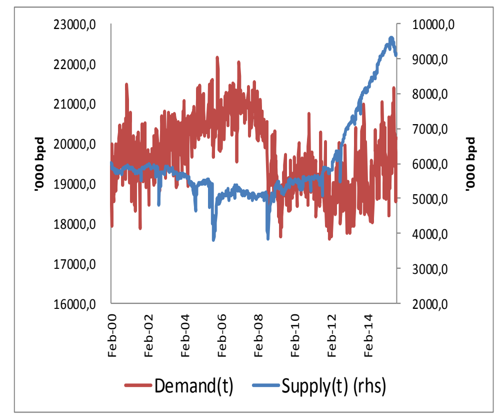
Update September 2015 (1/2)

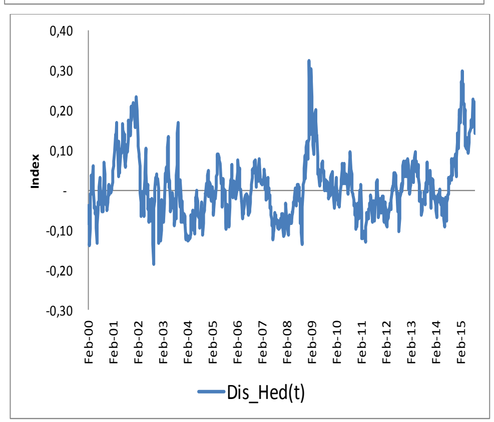
- Estimates are updated to include the more recent period (Dec 2014 – Sep 2015).
- Updated estimates fully include the recent price drop and surge in volatility.

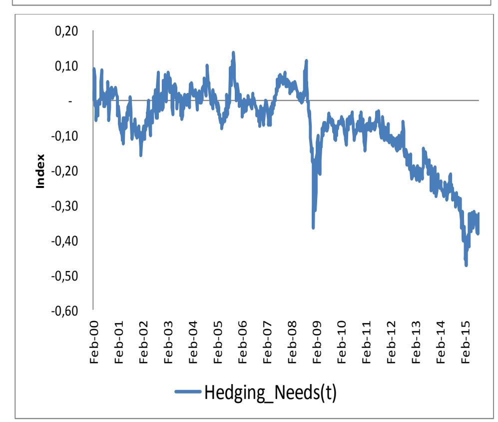




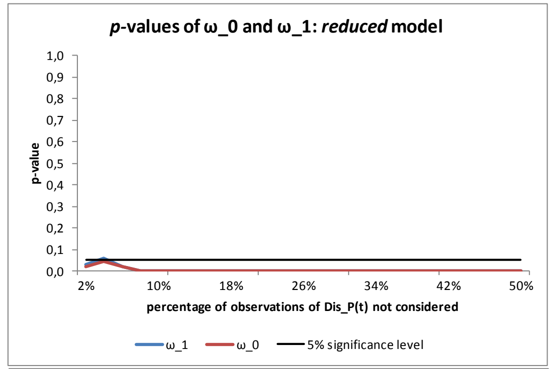


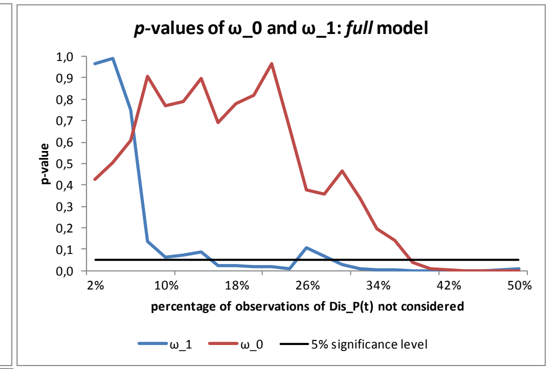






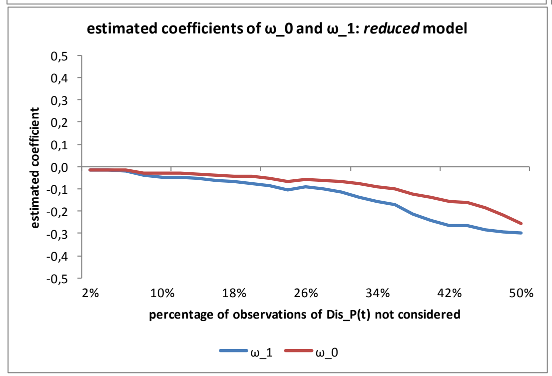
Update September 2015 (2/2)

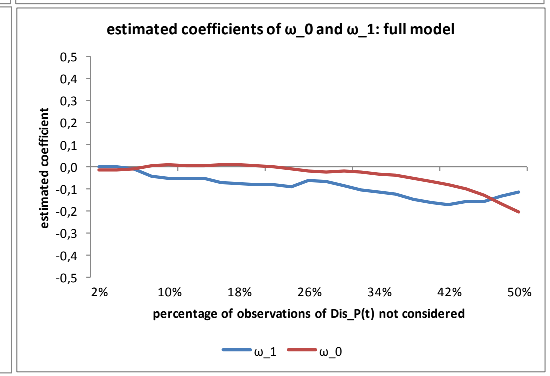


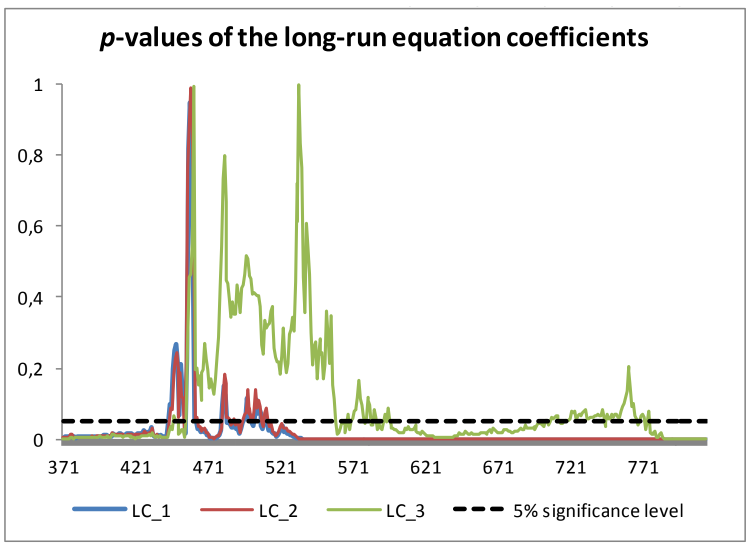


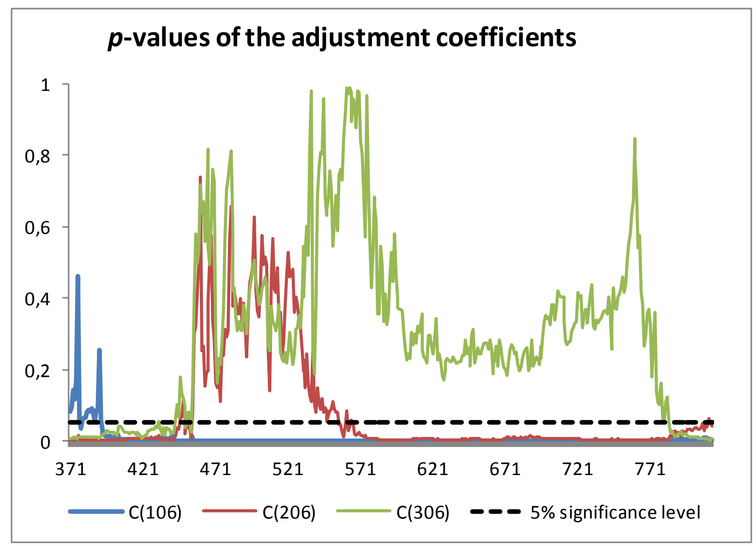
Empirical evidence shows:

- Similar results and coefficients significance of the price vs. speculation model.
- Hedging needs model still well performing.









THANK YOU

Elisa Scarpa
Elisa.scarpa@edison.it
Twitter: @ElisaScarpa

Paper joint with A. Cologni and F. G. Sitzia

Edison Trading S.p.A.
Foro Buonaparte, 31 - 20121 Milan – ITALY