

METODOLOGIA E PRIMI RISULTATI DI UN MODELLO PREVISIVO DEI PREZZI DEL PETROLIO

In this paper we analyze the main ingredients of a model employed to forecast the future price of the WTI listed on the Nymex. We present few results obtained from three different type of models: a single equation model, a multi-equational system and a VAR (Vector Autoregression) system. The forecasted values for the WTI futures front month are perfectly compatible with the values obtained from the forward curve of WTI.

Questo articolo analizza la metodologia seguita per la costruzione di un modello di previsione del prezzo del future WTI, prima scadenza, sul Nymex. Vengono passati in rassegna i criteri seguiti nell'impostazione di tre modelli previsivi: uniequazionale, multiequazionale simultaneo e VAR. I modelli forniscono previsioni sul prezzo future del WTI compatibili con l'andamento delle curve forward e con i dati effettivi.

La scelta di definire un modello di previsione del prezzo del petrolio può sembrare ad un primo tempo un azzardo. Gli scettici, infatti, diranno: «Ci hanno provato e ci provano in tanti e questo non è che l'ultimo tentativo». All'interno della società RIE, a scapito dello scetticismo, noi abbiamo dato vita all'impresa e i risultati – finora preliminari – sembrano essere altamente incoraggianti.

La crescita del grado di finanziarizzazione del mercato del petrolio – ci riferiamo al mercato del *West Texas Intermediate* (WTI), *light sweet crude oil*, quotato sul Nymex – ha creato l'esigenza di concentrare l'attenzione via via sul ruolo

degli strumenti finanziari che sono a disposizione degli operatori su questi mercati. In particolare, il mercato del petrolio è senza dubbio quello che evidenzia il maggior impatto della crescente finanziarizzazione dell'economia. Basti pensare, infatti, che tra tutte le transazioni regolate giornalmente sul Nymex riguardanti il WTI, soltanto lo 0,8% raggiunge l'effettiva regolazione fisica.

Tutto questo – evidentemente – pone problemi aggiuntivi riguardo alle necessità di considerare il prezzo del petrolio come se fosse il prezzo di una merce qualunque con l'esigenza di doverne prevedere anche gli andamenti futuri. Ciò è soprattutto urgente in un mercato come quello attuale in cui i punti di riferimento tradizionali sono saltati. La letteratura mostra, infatti, come il ruolo delle variabili reali si sia progressivamente attenuato nella sua capacità di spiegare l'andamento del prezzo del petrolio e dei suoi derivati. Al tempo stesso, tuttavia, la necessità di disporre di stime dei prezzi il più possibile attendibili è cresciuta quasi esponenzialmente: complice la sopraccitata finanziarizzazione che – inevitabilmente – ha comportato una crescita della volatilità.

Poter prevedere il percorso futuro del petrolio diviene assai rilevante anche in relazio-

* Università di Bologna,
Johns Hopkins University,
RIE S.r.l., Bologna
E-mail: marzo@economia.unibo.it

ne all'andamento dei prezzi dei prodotti derivati del petrolio (benzine, gasolio) o di prodotti ad esso correlati, quali ad esempio il gas naturale. Se poi si fa riferimento all'utilizzo del gas naturale come fonte privilegiata nella generazione di energia elettrica, ben si comprende quanto la previsione del prezzo del petrolio sia divenuta cruciale nel contesto attuale.

Abbiamo allora deciso di raccogliere la sfida e di iniziare un percorso verso la costruzione di un primo modello previsionale.

Va innanzitutto rilevato che la bontà di un modello previsionale dipende molto dalla qualità dei dati utilizzati nella fase di stima vera e propria del modello. In questa fase, abbiamo proceduto ad una serie di analisi preliminari dei dati, di rielaborazione, di costruzione/ricostruzione del campione e di ricerca di nuove fonti informative, al fine di arricchire lo stock di dati a nostra disposizione. Ciò come premessa anche di un futuro investimento verso l'acquisizione di database completi e organizzati, che comprendano informazioni non facilmente raggiungibili attraverso i tradizionali canali istituzionali.

È opportuno specificare fin da subito che la variabile oggetto della nostra previsione è rappresentata dal *future* prima posizione sul *West Texas Intermediate*, quotato sul Nymex. Per quale motivo ci siamo concentrati su questa variabile e non su altre? La prima spiegazione risiede nel fatto che il *future* WTI è la variabile più liquida riguardante il mercato del petrolio. Il volume delle transazioni giornaliera è molto elevato e l'*open interest* (lo stock di contratti attivi) molto corposo. Inoltre, il WTI risulta essere il tipo di greggio considerato come «guida» dagli

operatori, nel senso che le sue fluttuazioni di prezzo agiscono come «primo motore mobile» per contribuire all'analisi e spiegazione anche di altri greggi (ad es. il Brent) o altri prodotti derivati o correlati.

Il presente lavoro è strutturato come segue: dopo aver discusso le fasi che hanno condotto alla definizione del modello di previsione, si passerà ad analizzare il modello nel suo dettaglio per poi esaminarne gli utilizzi nonché una serie di sue estensioni.

1. IL LAVORO PRELIMINARE

L'elaborazione del modello è stata articolata su cinque fasi distinte: (1) reperimento dei dati; (2) analisi preliminare e trattamento dei dati; (3) stima di modelli preliminari; (4) selezione dei modelli e previsioni *in-sample*; (5) previsioni *out-of-sample*.

Come già detto, nella fase iniziale il nostro compito è consistito nell'assemblare più database provenienti da diverse fonti, provvedendo al reperimento anche di nuovi dati.

L'oggetto della previsione riguarda il WTI prima posizione, tuttavia, abbiamo ricostruito la serie storica giornaliera dei prezzi *futures* WTI per tutte le posizioni mediamente attive a partire dall'anno di inizio quotazione (1). Il nostro database di partenza ha frequenza giornaliera e prevede – oltre che i prezzi *futures* – anche informazioni relative all'*open interest* e i volumi trattati. Un database analogo è stato costruito e generato per la serie storica dei *futures* del gas naturale (Henry Hub), del gasolio da autotrazione e riscaldamento, della benzina senza piombo. Tali dati sono stati acquisiti presso il Nymex e successivamente rielaborati.

Per quanto riguarda l'insie-

me delle variabili di natura «reale», il nostro database comprende: l'ammontare delle scorte petrolifere, di carburante per aerei, nonché la domanda totale di prodotti raffinati, la sua offerta, le importazioni nette di petrolio (non distinte per area geografica di provenienza).

Altre variabili di natura reale o più squisitamente macroeconomica sono: indice della produzione industriale statunitense aggregato e settoriale, consumi (sia in aggregato sia distinti in beni durevoli e non durevoli), indici di fiducia del consumatore, l'insieme dei tassi che determinano la pendenza della *Term Structure* (2). Altre variabili del settore *energy* considerate nel modello sono i prezzi dell'olio combustibile atz e btz (cif e fob).

Dopo aver raccolto i dati da utilizzare come possibili ingredienti del modello, il passo successivo è consistito nell'analisi di ciascuna serie storica. In questo senso, abbiamo sottoposto le nostre variabili ad alcuni test di stazionarietà (3), in base ai quali possiamo concludere che quasi la totalità dei dati a nostra disposizione risulta essere non-stazionaria né in livelli né in logaritmi. Pertanto, per poter procedere ad una stima efficiente e puntuale (riducendo così il problema dell'autocorrelazione dei residui e dell'eteroschedasticità (4)) abbiamo scelto di utilizzare le variabili espresse in differenza prima (5) (evitando così di considerare le differenze prime logaritmiche, che configurerebbero un modello previsionale sui tassi di crescita delle variabili, con maggiori difficoltà analitico-interpretative).

Avendo realizzato che la gran parte delle variabili di interesse va espressa in termini di differenza prima, il passo successivo consiste nell'esaminare se ed in quale misura cia-

scuna variabile è in grado di fornire una previsione adeguata della differenza prima WTI. In pratica, questo secondo esercizio si sostanzia nel verificare quali variabili hanno la migliore capacità previsiva rispetto ad altre. A tale scopo, abbiamo utilizzato il test di Clark e McCracken (6). In senso lato, tale test ha come oggetto la valutazione della capacità previsiva di una variabile nei confronti di quella oggetto di interesse, nel nostro caso il *future* WTI prima scadenza. In altre parole, per selezionare le variabili da introdurre nel modello, abbiamo considerato ciascuna delle variabili del nostro database e abbiamo verificato se essa fosse in grado di fornire una buona previsione attendibile del WTI.

Poiché tale lavoro ha generato una pletora di risultati, in quanto segue ne selezioniamo solo alcuni. In primo luogo, tra le variabili che meglio contribuiscono a fornire una stima del WTI vi è la serie storica delle varie posizioni del *future* stesso (terza, sesta e nona), nonché alcuni prodotti, quali ad esempio il *future* su *heating oil* (prima scadenza) e su gas naturale (Henry Hub, prima scadenza). Inoltre, altre varia-

bili che dimostrano un discreto grado di capacità previsiva sono: l'indice della produzione industriale (generale e specifico), il *Federal Funds Rate*, il tasso di cambio euro/dollaro, il livello di fine periodo delle scorte di petrolio e altre variabili del settore *energy*.

Da queste riflessioni, ricaviamo che l'insieme delle variabili di cui sopra costituisce a pieno titolo un bacino da cui attingere per la costruzione del modello di previsione.

2. IL SETTING UP DEL MODELLO

Il passo successivo è rappresentato dalla definizione vera e propria del modello di previsione. A tale proposito, abbiamo ritenuto di procedere nell'elaborazione di tre modelli differenti (un quarto è in via di definizione). La motivazione principale nella definizione di un insieme di modelli così ampio risiede prevalentemente nella necessità di porre in competizione i modelli l'uno con l'altro, al fine di individuare quello con la miglior capacità previsiva. Inoltre, alcuni modelli catturano meglio una dinamica di lungo periodo, altri una dinamica di breve/medio periodo. Le previsioni risultanti da ciascun modello sono tutte utilizzate per definire un'unica previsione, da presentarsi all'esterno.

I modelli finora considerati sono: (1) un modello uniequazionale; (2) un modello a equazioni simultanee; (3) un modello VAR (*vector autoregressions*).

Modello uniequazionale. In questo caso il *future* WTI è funzione di una serie di variabili, tra le quali abbiamo il *future* WTI stesso (ritardato di un periodo); la terza posizione del *future* WTI (contemporanea e ritardata); lo stock di riserve di

petrolio negli Stati Uniti, l'indice di produzione industriale, il *Federal Funds Rate* (FFR), il tasso di cambio euro/dollaro e il *future* su *heating oil* ritardato di un periodo. Il modello ha un'ottima capacità previsiva *in-sample* (7) e una discreta capacità previsiva *out-of-sample* (8). Tuttavia, qual è il problema di questo approccio? Il problema fondamentale risiede nel fatto che le variabili da cui dipende il *future* WTI sono da considerarsi esogene. Pertanto, nell'elaborazione della stima *out-of-sample* è necessario determinare ex ante il sentiero di movimento di tali variabili, che – evidentemente – è formulato in maniera abbastanza arbitraria. Per questo motivo, abbiamo ritenuto opportuno utilizzare il modello a equazioni simultanee.

Modello a equazioni simultanee. Tale modello considera un insieme di equazioni che «spiegano» una serie di variabili le quali, a loro volta, possono entrare o meno nella definizione del *future* WTI oggetto delle previsioni. Il modello che abbiamo considerato comprende sette equazioni composte da due blocchi distinti:

- variabili endogene: gas naturale, terza posizione WTI, prima posizione WTI, *Heating Oil futures* prima posizione;
- variabili esogene: FFR, produzione industriale, tasso di cambio.

Il modello è stimato attraverso diversi metodi (9). Il confronto con le stime ottenute in base ai vari metodi non rivela particolari differenze, segnale che il modello è ben specificato. Per la definizione delle previsioni abbiamo considerato come base il metodo SURE, in quanto ci consegna una stima migliore della matrice varianza/covarianza. Alla base di tutte le stime (a parte la FIML) vi è l'utilizzo di una serie di stru-



menti che consentono di rendere il procedimento di stima più corretto e consistente. Tra le variabili utilizzate come strumenti vi sono: WTI sesta posizione; *open interest* WTI prima posizione; *capital utilization rate*; *consumer confidence index*; scorte petrolifere. Il vantaggio principale del modello a equazioni simultanee sta tutto nella sua capacità di sfruttare al massimo l'informazione che deriva dalla stima simultanea e dall'interazione delle variabili del modello.

Riguardo alla previsione *in-sample* i risultati sono molto incoraggianti. Per quanto riguarda la stima *out-of-sample*, invece, i risultati sono raccolti nella Tab. 1, da cui emerge che il modello ha un'ottima performance se confrontato con i dati reali. Infatti, il modello utilizzato per le previsioni *out-of-sample* è stato stimato fino al settembre 2006: la prima previsione riguarda dunque ottobre-dicembre 2006 e gennaio-aprile 2007. Il nostro modello prevede una leggera crescita del prezzo rispetto ai valori medi di ottobre, lungo il corso del 2007. Per la fine del 2006, si prevede una sostanziale tenuta dei prezzi, senza eccessive oscillazioni. Questo risultato appare del tutto compatibile con l'andamento delle curve *forward* che prevedono per i mesi futuri una sostanziale tenuta dei prezzi, con un leggero rialzo (dovuto prevalentemente al ben noto fenomeno di *backwardation*) da registrarsi nel corso dei primi mesi del 2007. I valori presentati in Tab. 1 sono coerenti con le previsioni offerte sul mercato da altri *forecasters* e dall'Agenzia Internazionale per l'Energia. Riteniamo pertanto tali risultati assolutamente confortanti per il successivo sviluppo del modello.

Il modello VAR. L'ultima categoria di modelli considerata nelle nostre stime è rappre-

Tab. 1 - PREVISIONI DERIVANTI DAL MODELLO A EQUAZIONI SIMULTANEE

	West Texas Intermediate		Henry Hub natural gas
	prima scadenza doll./bbl	terzo scadenza doll./bbl	prima scadenza doll./mil. btu
2006			
Giugno	70,969	72,372	6,385
Luglio	74,436	76,387	6,216
Agosto	73,083	75,107	6,989
Settembre	63,895	65,750	5,218
Ottobre	60,315	62,331	4,808
Novembre	60,800	62,717	4,673
Dicembre	60,564	62,530	4,686
2007			
Gennaio	60,956	62,901	4,688
Febbraio	61,090	63,050	4,708
Marzo	61,355	63,313	4,722
Aprile	61,555	63,519	4,739

sentata dal modello *Vector Autoregression* (VAR), largamente utilizzato in macroeconometria come modello di medio termine per la sua capacità di prevedere – su dati trimestrali – le grandezze economiche a 4-5 trimestri in avanti. Il vantaggio di questo modello risiede nella grande parsimonia di variabili utilizzate, dal momento che è la loro interazione a vari lag temporali che attribuisce a questa categoria di modelli una ottima capacità previsiva.

L'insieme delle variabili da noi considerato per la stima del VAR è il seguente: *future* WTI prima, terza e nona scadenza; *future* gas naturale (prima scadenza). Nei modelli VAR l'identificazione gioca il ruolo principale, dal momento che definisce l'insieme dei legami causali del modello. Come prima

approssimazione, abbiamo scelto un criterio di identificazione ricorsivo a blocchi (*block recursive*) in cui il prezzo del WTI prima scadenza influenza tutte le altre variabili all'interno dello stesso periodo iniziale t , ma non ne è da esse influenzato. I risultati della previsione *out-of-sample* sono riportati in Tab. 2: il modello fornisce una previsione più «rialzista» rispetto al precedente, anche se il trend è pressoché identico. La tendenza rialzista è anche confermata nella parte di previsioni *in-sample*: essa gioca come un fattore di scala che, se opportunamente considerato, consente di riportare il modello sugli stessi valori ottenuti in precedenza.

Dal confronto con i risultati di Tab. 1 ricaviamo che i modelli si complementano a vicenda. Tuttavia, anche se i ri-

Tab. 2 - PREVISIONI DERIVANTI DAL MODELLO VAR

	West Texas Intermediate		Henry Hub natural gas
	prima scadenza doll./bbl	terza scadenza doll./bbl	prima scadenza doll./mil. btu
2006			
Giugno	70,969	72,372	6,385
Luglio	74,436	76,387	6,216
Agosto	73,083	75,107	6,989
Settembre	63,895	65,750	5,218
Ottobre	61,326	63,404	4,782
Novembre	63,484	65,749	4,716
Dicembre	64,245	66,437	5,178
2007			
Gennaio	63,551	65,498	5,508
Febbraio	62,410	64,226	5,745
Marzo	61,714	63,493	5,926
Aprile	61,349	63,111	6,078

sultati ora presentati appaiono abbastanza soddisfacenti, il nostro obiettivo è di affinare tali modelli, introducendo altre variabili o sostituendo quelle già presenti con altre, o modificando l'identificazione o i metodi di stima (come nel caso del VAR ad esempio). Un'altra strada interessante da percorrere (che, peraltro, stiamo già affrontando), riguarda la costruzione di un modello VECM (*Vector Error Correction Mechanism*), ovvero un modello vettoriale a correzione dell'errore. In qualche modo, questo modello può essere pensato come un compromesso tra il modello a equazioni simultanee e il modello VAR, dal momento che congiunge un sistema di relazioni stabili di lungo periodo tra alcune variabili di interesse (ad esempio il legame tra le varie posizioni del *future* WTI, già identificato dal modello VAR) con quelle derivanti dall'interazione con un set di variabili esogene.

3. UTILIZZI ED ESTENSIONI

Del modello di previsione abbiamo pensato ad alcuni utilizzi principali, ad una estensione, ma soprattutto esso ci è di stimolo per una riflessione più profonda sugli strumenti da impiegare nell'analisi del mercato dell'energia.

Come possibile estensione il primo passo da fare a seguito di tale analisi consiste nell'effettuare una serie di previsio-

ni, per così dire «a cascata». Ovvero, a partire dalla previsione generata sul *future* WTI possiamo generare una serie di previsioni per i prodotti strettamente correlati ad esso: il Brent, il gas naturale nel Regno Unito, olio combustibile atz e btz e, per conseguenza, i prodotti in Europa. Tali previsioni sono realizzate utilizzando la matrice di correlazioni contemporanee e *forward* esistente tra il *future* WTI sul Nymex e il resto delle variabili di interesse. In questo modo, otteniamo una serie di previsioni anche su un insieme di variabili di grande interesse (comprendente anche i prezzi cif e fob del gas naturale quotato nel Regno Unito, a Zeebrugge e a Bunde) per le applicazioni pratiche.

Una linea di utilizzo della previsione riguarda la possibilità di esplorare più da vicino il ruolo di alcuni mercati collegati al WTI: ci riferiamo in particolare al mercato del gas naturale e a quello degli altri greggi, quali l'Arabian Light, i greggi del Far East e altri ancora. Evidentemente, si tratterebbe di analisi che la semplice struttura di correlazione sopra descritta non catturerebbe pienamente.

In ogni caso, lo sforzo che abbiamo posto in essere nell'analisi e nell'elaborazione del modello di previsione ci porta ad una riflessione più ampia sulla metodologia adottata nelle analisi del mercato dell'energia. È evidente che, alla luce

dei processi che caratterizzano tale mercato da qualche anno a questa parte, schemi di analisi esclusivamente basati sulla realtà fattuale tendono a fornire una rappresentazione non completa dei fenomeni in atto. Si sottolinea: una rappresentazione non completa, ma non per questo errata. Crediamo che – sulla scorta anche delle tendenze di altri centri di ricerca a rilevanza nazionale ed internazionale – sia opportuno un salto di qualità nell'analisi delle dinamiche di prezzo del mercato petrolifero ed energetico in generale. È allora in questa ottica che abbiamo desiderato investire sul modello di previsione per avviare un percorso di intenso monitoraggio del mondo dei derivati finanziari legati al mondo dell'energia (che comprendono anche, tra gli altri, i *green certificates*).

Naturalmente il modello di previsione presentato è suscettibile di miglioramenti, estensioni ed integrazioni sia sul piano della metodologia, sia su quello della base dati utilizzata ai fini della stima, che deve essere costantemente aggiornata e ampliata. L'attività di previsione si sostanzia nella capacità di interagire con il mercato, con gli operatori al fine di confrontare la *view* che proviene dal modello con il *sentiment* degli operatori, sia ex-ante che ex-post la presentazione della stima.

Bologna, Novembre 2006

Ringrazio il Prof. Alberto Clò per i continui stimoli ed incoraggiamenti. Stefano Verde e Lisa Orlandi hanno contribuito in maniera cruciale alla definizione del database: a loro va il mio più sincero ringraziamento. Un grazie anche a Patrizia Bassani, per la pazienza e competenza dimostrata nel corso della stesura del lavoro. Nessuna delle persone nominate è responsabile per errori o manchevolezze, che sono da addebitare unicamente all'autore.

NOTE

(¹) Non tutte le posizioni del WTI sono state attivate fin dal giorno dell'inaugurazione del contratto *future*. Per questo motivo, il database completo di tutte le posizioni inizia a partire dal gennaio 1987.

(²) La *Term Structure* (ovvero la relazione tra i rendimenti dei titoli di Stato di differente durata e la loro scadenza) è data, nel nostro caso, dall'insieme dei seguenti tassi: *Federal Funds Rate*, tasso dei titoli di Stato a 3 mesi, a 10 e a 20 anni; stock di M1 ed M2, indice Dow Jones, Standard and Poor's 500, tasso di cambio euro/dollaro.

(³) In termini molto grezzi, un processo stocastico si definisce stazionario se i suoi parametri non dipendono dal tempo. I test di stazionarietà qui utilizzati sono: Augmented Dickey-Fuller, Ng and Perron, Phillips and Perron.

(⁴) In statistica, e in particolare nell'ambito dell'econometria, si parla di eteroschedasticità allorché i residui otte-

nuti in seguito ad un procedimento di stima presentano caratteristiche di (1) non-stazionarietà e (2) una varianza continuamente mutevole nel corso del tempo.

(⁵) Per differenza prima si intende la differenza tra la variabile al tempo t e la medesima variabile al tempo $t - 1$.

(⁶) CLARK T.E., McCracken M.W. (2001), *Test of Equal Forecast Accuracy and Encompassing for Nested Models*, in «Journal of Econometrics», vol. 105, pp. 85-110.

(⁷) I grafici e la diagnostica della stima dell'equazione sono disponibili su richiesta.

(⁸) Si definiscono previsioni *out-of-sample* quelle ottenute al di fuori del campione considerato per la stima.

(⁹) I metodi di stima utilizzati sono: *Two Stage Least Squares* (2SLS), *Three Stage Least Squares* (3SLS), *Full Information Maximum Likelihood* (FIML), *Seemingly Unrelated Regressions* (SURE).