Energy Economics 32 (2010) 409–417

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Macroeconomic factors and oil futures prices: A data-rich model☆

Paolo Zagaglia

Modelling Division, Sveriges Riksbank, Sweden

ARTICLE INFO

Article history:
Received 26 March 2009
Received in revised form 8 November 2009
Accepted 8 November 2009
Available online 14 November 2009

JEL classification:
C53
E52

Keywords:
Crude oil
Futures markets
Factor models

ABSTRACT

I study the dynamics of oil futures prices in the NYMEX using a large panel dataset that includes global macroeconomic indicators, financial market indices, quantities and prices of energy products. I extract common factors from the panel data series and estimate a Factor-Augmented Vector Autoregression for the maturity structure of oil futures prices. I find that latent factors generate information that, once combined with that of the yields, improves the forecasting performance for oil prices. Furthermore, I show that a factor correlated to purely financial developments contributes to the model performance, in addition to factors related to energy quantities and prices.

1. Introduction

During the past year, oil prices have made the headlines of the financial press almost every day. Since the beginning 2008, the spot price of crude oil traded in the New York Mercantile Exchange (NYMEX) has almost doubled at peak. This has raised serious concerns among market participants and policymakers worldwide. Comments released to the press have often denoted a deep disagreement on the causes of the price spikes and, in general, on the mechanics of oil market.

Bernanke (2008) has represented the central bankers' view in a timely manner, stating that[1]

> "…the price of oil has risen significantly in terms of all major currencies, suggesting that factors other than the dollar, notably shifts in the underlying global demand for and supply of oil, have been the principal drivers of the increase in prices. (…) Another concern that has been raised is that financial speculation has added markedly to upward pressures on oil prices. (…) However, if financial speculation were pushing oil prices above the levels consistent with the fundamentals of supply and demand, we would expect inventories of crude oil and petroleum products to increase as supply rose and demand fell. But in fact, available data on oil inventories show notable declines over the past year."

Since oil commodities are traded through futures and derivatives contracts, market views shape the pricing of oil commodities. In this sense, the financial press has pushed the hypothesis that purely 'financial' considerations, unrelated to 'real' market developments, have been behind the recent spikes (see Chung, 2008 and Mackintosh, 2008).

The distinction between financial and real determinants of oil prices in the long run is also present in the academic literature. A large number of papers suggest that oil prices are mainly driven by demand and supply considerations. For instance, Kilian (2008b) suggests that a proper measurement of the business cycle effects of energy prices requires disentangling the role of demand supply shocks in energy markets. Kilian (2008a) decomposes the real price of crude oil into supply shocks, shocks to the global demand for industrial commodities, and demand shocks that are idiosyncratic to the oil market. The role of energy quantity factors is stressed also in Alquist and Kilian (2008), who show that spread between oil futures prices of different maturities are related to uncertainty about supply shortfalls.

The literature on the financial determinants of oil prices has produced various contributions on the role of market uncertainty and volatility for oil pricing. Askari and Krichene (2008) model the jump intensity of daily crude oil prices between 2002 and 2006. They find that measures of market expectations extracted from call and put option prices have incorporated no change in underlying fundamentals in the

☆ This work started when I was a visiting scholar in the Research Unit of the Bank of Finland, whose hospitality is greatly acknowledged. I am deeply grateful to the editor (Richard Tol) and to two anonymous referees for very thorough and helpful comments. I also acknowledge the assistance of Vincent Brousseau and Tarja Yrjölä in the construction of the dataset. Efrem Castelnuovo, Juha Kilponen, Massimiliano Marzo and Alistair Milne contributed with relevant suggestions and constructive criticism. The views expressed here are my own and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank. This research was carried out by the author for purely academic purposes and without financial support from any private institution.

E-mail address: paolo.zagaglia@gmail.com.
URL: http://pz.pcriot.com.

[1] A similar argument is discussed by Trichet (2008).

0140-9883/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.eneco.2009.11.003

short term. Chong and Miffre (2006) document the presence of a significant pattern of risk premia earned by investors on a number of commodities futures since 1979, including crude oil. Gorton et al. (2007) show that, although commercial positions on oil futures are correlate with inventory signals, they do not determine risk premia.

The presence of two opposing views on price formation in the oil market over the long run implies that a number of key questions are not dealt with in the literature. The issue of causality between spot and futures prices across the maturity structure is largely unsettled. Suppose that oil futures contain information about spot prices. Omitting futures prices would bias the results in favour of a strong role for demand–supply factors to drive the spot price. Moreover, the role of macroeconomic factors for the dynamics of oil prices is typically studied in isolation from the conditions prevailing in financial markets.[2]

In this paper I study whether the interplay between real and financial factors can play a systematic role for explaining oil prices changes over a long time period. I exploit the information from a large panel to investigate the sources of changes in the term structure of futures prices for WTI crude oil. Like Bernanke et al. (2008), I extract common factors form the large panel dataset, and I model the joint dynamics of the factors and the oil prices in a 'Factor-Augmented' Vector Autoregression (FAVAR). The factors mimic the drivers of oil prices that are 'latent,' in the sense that they are not directly observed by the econometrician from the information set. In standard Vector Autoregressive (VAR) models, the econometrician is required to choose what observable variables best represent theoretical concepts, such as supply and demand. The supply of oil can be measured with data on oil production. However, these data series are affected by measurement errors of different types, for instance arising from aggregation. As argued by Bernanke et al. (2008), the use of sparse information in the form of factors extracted from a large dataset mitigates this problem.

This modelling strategy has already been applied by Ludvigson and Ng (forthcoming) and Mönch (2008) for the construction of pricing models for the yield curve of government bonds, and it presents several advantages. The model can capture the interdependence between oil price changes and the factors of different nature. The FAVAR also allows to model jointly the relevant maturities of oil futures prices in a flexible way. It should be stressed that the literature features a long list of contributions on the role of unobservable factors for oil price dynamics.[3] These contributions differ from the present paper in two dimensions. The factors are typically meant to drive the time-varying volatility of observed at a daily frequency. Instead here I use monthly data, and I abstract from the role of high-frequency price movements.

The panel dataset from which I extract common components include over 200 data series with detailed information on energy demand and supply, energy prices, macroeconomic and financial variables. I show that a latent factor correlated with the open interest on oil futures prices contributes significantly to the joint model of the oil price returns. This appears to corroborate the conjecture of Trichet (2008) on the financial determinants of oil prices. The other factors are strongly correlated with data on energy quantity and prices, as typically suggested by the macroeconomics literature. I find that augmenting the information from the term structure of oil futures prices with latent factors improves the forecasting performance of the model.

[2] For instance, Trichet (2008) suggests that the combination of real and financial factors have played a joint role in the recent episodes of oil price movements.

[3] A non-exhausting list consists of Casassus and Collin-Dufresne (2005), Carmona and Ludkovski (2005), Gibson and Schwartz (1990), Pindyck (2004), Postali and Picchetti (2006), Radchenko (2007), Schwartz (1997), and Schwartz and Smith (2000).

This paper is organized as follows. In Section 2, I outline the structure of the FAVAR model. Section 3 presents the dataset. Section 4 describes the results. Section 5 concludes.

2. The factor-augmented VAR model

The model presented here is based on the assumption that the futures price for one maturity is driven both by the prices of the other maturities, and by macroeconomic shocks. The macroeconomic determinants are proxied by unobservable factors that summarize the common information in a large number of time series. The joint dynamics of the observable an unobservable variables in modelled in the FAVAR model of Bernanke et al. (2008).

The general form of the FAVAR can be written as

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \mu + \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \nu_t \tag{1}$$

where $\Phi(L)$ is a matrix of lag polynomials, and ν_t is a vector of normally-distributed shocks. Y_t is a vector $m \times 1$ of observed variables. The unobservable factors are collected in the $k \times 1$ vector F_t. Eq. (1) states that the dynamics of the factors is affected by its own lags, by the vector of observables, and by the shocks. The model 1 has a variance–covariance matrix Σ.

Eq. (1) cannot be estimated without knowledge of F_t. For that purpose, a large number p of series can be used to extract 'common' factors. The 'information series' are collected in the vector X_t with dimension $p \times 1$. The dynamic factor model of Stock and Watson (2002) can then be used to obtain F_t. This framework assumes that the information time-series X_t are related to the factors F_t and the observed variables Y_t through the observation equation

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + \epsilon_t. \tag{2}$$

where Λ^F is a $p \times k$ matrix of factor loadings. The measurement equation formalizes the idea that both the oil futures returns and the factors drive the dynamics of the panel dataset. In other words, the factors can be measured with noise from the panel dataset.

Bernanke et al. (2008) propose two methods for estimating the model 1–2. The first one is the 'diffusion index' approach of Stock and Watson (2002), which consists itself of two steps. In the first step, Eq. (2) is used to estimate the unobservable factors F_t through asymptotic principal components. Then, the estimated factor $\hat{F}_t$ is fit to the FAVAR model 1. The second estimation method follows a single-step Bayesian likelihood approach. Bernanke et al. (2008) discuss a Gibbs sampler that approximates the marginal posterior densities of both the factors and the parameters. Since it is not clear a priori which estimation method delivers the results that are most desirable, Bernanke et al. (2008) estimate the model using both approaches, and find that they yield similar outcomes.

In this paper, I apply the two-step estimation procedure. Thus, I extract unobservable factors by using the asymptotic principal component method. Extracting the common factors from the panel dataset consists in recovering the space of X_t spanned by F_t. Denote by V the eigenvectors corresponding to the k largest eigenvalues of the variance–covariance matrix XX'/k. The estimated factors are $\hat{F} = \sqrt{T}V$, where T is the time-series length, and the factor loadings are $\hat{\Lambda}^f = \sqrt{T}X'V$.

3. The dataset

I use monthly data from January 1992 until March 2008 for a total of 193 observations for each series. The vector Y_t consists of returns on the spot price for WTI crude oil traded in the New York Mercantile Exchange (NYMEX), and on futures prices with maturities of 1, 6 and