

# Estimating Oil Risk Factors Using Information from Equity and Derivatives Markets

I-HSUAN ETHAN CHIANG, W. KEENER HUGHEN, and JACOB S. SAGI\*

November 26, 2013

---

\*Ethan Chiang and Keener Hughen are from the Belk College of Business, University of North Carolina at Charlotte. Jacob Sagi is from the Kenan-Flagler Business School, University of North Carolina at Chapel Hill. Corresponding author: Keener Hughen, Belk College of Business, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223; Email: [whughen@uncc.edu](mailto:whughen@uncc.edu). Ethan Chiang can be reached via [ichiang1@uncc.edu](mailto:ichiang1@uncc.edu); Jacob Sagi can be reached via [sagij@kenan-flagler.unc.edu](mailto:sagij@kenan-flagler.unc.edu). We are grateful to seminar participants at the University of Colorado, the University of Florida, the University of Maryland, and the 25th Annual Conference of the Financial Markets Research Center at Vanderbilt University for their useful comments. We particularly benefitted from comments by Gurdip Bakshi, Bernard Dumas, Pete Kyle, Chris Leach, Andy Naranjo, Sugata Ray, Georgios Skoulakis, and Zijun Wang. Jun Chen provided excellent research assistance. Ethan Chiang and Keener Hughen acknowledge the support of the Belk College Summer Research Grant.

# Estimating Oil Risk Factors Using Information from Equity and Derivatives Markets

## ABSTRACT

We introduce a novel approach to estimating latent oil risk factors and establish their significance in pricing non-oil securities. Our model, which features four factors with simple economic interpretations, is estimated using both derivative *prices* and oil-related equity returns. The fit is excellent in and out of sample. The extracted oil factors carry significant risk premia, and are significantly related to macroeconomic variables as well as portfolio returns sorted on characteristics and industry. The average non-oil portfolio exhibits a sensitivity to the oil factors amounting to a sixth (in magnitude) of that of the oil industry itself.

**JEL:** G12, G13

**Keywords:** Oil, futures, asset pricing, APT, real options, CAPM, Fama-French factors, Markov chain

Monte Carlo

Few if any commodities have been the focus of more attention for their perceived economic significance than oil. While there is strong evidence relating oil prices to the business cycle, the nature of the relationship is nonlinear, time-varying, and difficult to attribute to any single source such as political uncertainty, cartel decisions, or global economic conditions (see Hamilton, 2003; and Barsky and Kilian, 2004). Despite its prominence in the business media and economics literature, and despite the well-documented role of business cycles in asset pricing, academic research has largely failed to find consistent evidence that oil is an important determinant of cross-sectional asset prices.<sup>1</sup> This paper introduces a new model and method to estimate latent oil risk factors using noisy price information from equity and derivative markets simultaneously. Put simply, we extract four distinct oil price components and find that these are important in explaining the movement of asset prices and macroeconomic fundamentals. While similar models and techniques have in the past been used to explore the valuation of commodity derivatives, we are the first to tie the methodology to the pricing of equities.<sup>2</sup> In particular, the studies that found significant risk premia in oil-related factors (e.g., Trolle and Schwartz, 2009; Casassus and Collin-Dufresne, 2005) have not established whether these risk premia are economically relevant outside oil derivative markets. The oil factors we extract appear to have an economically and statistically significant relationship with macroeconomic variables such as real GDP, industrial production, unemployment, inflation, and market uncertainty (the VIX). Our four oil factors explain 23.6% percent of the variance of oil industry returns beyond what can be explained using market returns alone, and this oil risk exposure is associated with a risk premium of 5.3%. The magnitude of oil risk exposure exhibited by non-oil characteristic and industry portfolios is on average about a sixth of that of the oil industry, and compensated by a commensurate risk premium. Overall, our work can be seen as a new approach to Ross' (1976) Arbitrage Pricing Theory, employing structural models together with data from several linked markets to extract a less noisy set of fundamental and systematic pricing factors.

In our econometric specification, the log-price of oil is affected by four distinct components.

One can be thought of as transient while another may be viewed as persistent. The drift in the persistent component is itself a factor, reflecting the long-run trend in oil prices. Finally, the fourth component drives the stochastic volatility of oil prices (known to exhibit heteroskedasticity). Thus each of our model's components has a ready interpretation in terms of fundamental news affecting oil prices, which is potentially important in sorting out how oil shocks impact the macroeconomy (see Hamilton, 2003; and Barsky and Kilian, 2004). We estimate the model using oil futures prices and option-implied variance, augmented with returns on oil-related stocks. The latter can be an important source of information about oil prices that may not be captured by short-dated derivatives. The model is estimated using 30 years of daily data on futures, options, and equity returns, fitting derivative prices and returns very well both in and out of sample. The latent factors identified by our estimation methodology carry significant risk premia, both statistically and economically, with magnitudes of unconditional annual Sharpe Ratios as high as 0.43. Of the four, the oil volatility factor appears to be most related to non-oil variables, including the macroeconomic variables mentioned earlier as well as the Fama-French factors.

The paper makes several contributions to the literature. First, we demonstrate that stock prices contain important information about fundamentals, such as commodity prices, that is not contained in commodity derivatives. By combining information from both markets we succeed in explaining oil-related security prices and returns better than traditional asset-pricing approaches. Second, our model and the estimation methodology are important to the real options literature where valuation relies disproportionately on an accurate description of persistent dynamics and stochastic volatility. Third, we demonstrate that oil shocks are systematic and command non-trivial prices of risk. Finally, among the types of news about oil, our approach identifies shocks affecting the magnitude of uncertainty in oil prices as most important to the macroeconomy and cross section of expected returns.

We do not presume that oil (combined with the market) spans all pricing factors. The paper seeks to provide insights to anyone who believes that oil plays an important role in the pricing of

securities — regardless whether one is motivated by ad-hoc or general equilibrium considerations to include oil in an empirical asset-pricing model. Our study suggests that stochastic oil volatility is as important, if not more so, for the macroeconomy than oil price itself. It also suggests that permanent and temporary oil price shocks have different effects on asset prices. In that sense, the paper is best viewed as a thorough attempt to study how oil risks are linked to oil and non-oil securities rather than as a pre-specified factor model like that of Chen, Roll, and Ross (1986). Our approach and findings complement recent work that explicitly models the role of oil in the pricing of securities (see Baker and Routledge, 2012; and Ready, 2012, 2013).

The paper is structured as follows. Section I introduces our four-factor model, describing its dynamics, derivative prices, and the link between equity returns and the model risk factors. Section II describes the data we employ and the estimation methodology (Markov chain Monte Carlo). Section III presents the estimation results and the cross-sectional analysis of the extracted latent factors. The last section concludes. The Appendix contains details of the model derivation and estimation procedure. A supplementary Internet Appendix contains further information.

## I. Model

Our goal is to estimate oil-relevant state variables from observed prices of oil futures, options, and equities. To do so efficiently, we look for a specification admitting closed-form solutions for futures and option prices in terms of fundamental state variables. Following the literature on term-structure modeling for interest rates and commodities we opt for an affine specification (see, for example, the references in Collin-Dufresne, Goldstein, and Jones, 2009). Intuitively, one should be able to distinguish three types of factors that influence oil prices and their forecasts: a transient factor, a growth-trend (i.e., long-run forecast) factor, and a permanent shock factor. All three factors should be reflected in futures prices through the near-term slope, long-term slope, and level of the term structure. Moreover, Trolle and Schwartz (2009) and Hughen (2010) demonstrated the presence of a fourth factor, not spanned

by futures prices, that impacts volatility. Thus, a minimal realistic description suggests the use of a four-factor affine model given by the following risk-neutral Itô diffusion:

$$dX_t = (A^* + B^* X_t) dt + \Sigma_t dW^*,$$

where  $X_t$  is a four-dimensional state variable and  $\Sigma_t \Sigma_t'$  is affine in  $X_t$ . In Appendix A we derive the parametrizations of the coefficient matrices  $A^*$ ,  $B^*$ , and  $\Sigma_t \Sigma_t'$  implied by the presence of a persistent component, a single and unspanned volatility component, and identifiability. The spot price of oil evolves as  $\log S_t = X_{s,t} + X_{p,t}$ , and as will be justified shortly, the vector of oil factors,  $X_t' = (X_{s,t}, X_{l,t}, X_{p,t}, X_{v,t})$  corresponds, respectively, to the transient (short-term), growth-trend (long-term), persistent, and unspanned volatility factors.

Following standard practice, the dynamics of  $X_t$  under the physical measure are also assumed to have an affine drift of the form:  $A + BX_t$ . The first three elements in the fourth row of the matrix  $B$  must be zero if the instantaneous variance is to be positive definite under the physical measure. The instantaneous risk premium associated with the oil factors is the difference between the physical and risk-neutral drifts,

$$\Lambda_t = A - A^* + (B - B^*)X_t. \tag{1}$$

Constraining the risk premium to be stationary restricts the third column (corresponding to the persistent variable  $X_{p,t}$ ) of the matrix  $B$  to be zero.<sup>3</sup> We note that the risk premium is dynamic, consistent with recent general equilibrium models of oil price risk.<sup>4</sup>

**Futures** The futures price  $F_{t,\tau}$  at time  $t$  for delivery at time  $t + \tau$  of one unit of oil is derived in Appendix B and given by

$$F_{t,\tau} = e^{\chi(\tau) + \eta_s(\tau)X_{s,t} + \eta_l(\tau)X_{l,t} + \eta_p(\tau)X_{p,t}}, \tag{2}$$

where  $\eta_s(\tau)$  declines with maturity,  $\eta_l(\tau)$  increases with maturity, and  $\eta_p(\tau)$  is constant. Because  $\eta_s$  declines with maturity, an increase in  $X_{s,t}$  will increase the short end of the log-futures curve but will have little effect on the long end. Intuitively, a shock to  $X_{s,t}$  may correspond to changes in oil prices that will quickly induce high-cost producers to enter or exit the market, or cause consumers to switch to readily available energy alternatives. By contrast, an increase in  $X_{l,t}$  will have little effect on the short end but will increase the long end, reflecting news about the long-run trend to which producers and consumers cannot (or need not) readily adjust. Finally, an increase in  $X_{p,t}$  will shift the entire log-futures curve upward by the same amount across all maturities, representing a permanent shock and consistent with the fact that oil is an exhaustible resource and sensitive to technological shocks. This confirms the postulated roles of the three risk factors and allows us to identify them with different types of economic news concerning oil prices. Note that, by definition, futures prices do not explicitly depend on the unspanned volatility state variable,  $X_{v,t}$ .

Much of the literature on commodity derivative prices focuses on explicitly modeling the net convenience yield, which represents the benefits foregone by those who have forward rather than current rights to the use of a commodity. In an Itô diffusion setting, the instantaneous net convenience yield,  $\delta_t$ , is defined by the following risk-neutral evolution:

$$\frac{dS_t}{S_t} = (r_t - \delta_t) dt + \Sigma_t dW^*,$$

where  $r_t$  is the prevailing instantaneous risk-free rate. In our setting, one can write the convenience yield as

$$\delta_t = r_t + \kappa_s X_{s,t} - X_{l,t}$$

where  $\kappa_s$  is the speed of mean reversion of  $X_{s,t}$ . To interpret this equation, consider that a temporary increase in oil price (i.e.,  $X_{s,t}$  increases) benefits those who have oil in inventory more than those who have future claims to oil. The more transient the shock (i.e., the larger  $\kappa_s$  is), the greater the relative

benefit to inventory owners. Likewise, an increase in the future price of oil (i.e.,  $X_{l,t}$ ) benefits forward claim holders more than inventory holders because the latter have to pay storage costs to cash in on the higher future price of oil. Finally, if interest rates are high while oil prices are expected to remain constant, anyone in current possession of oil can sell it and earn a high interest rate on the proceeds, only to buy the oil back in the future if and when it is needed.

**Options** The CME lists options on crude oil futures contracts from which one can hope to extract information about the unspanned volatility state variable,  $X_{v,t}$ . Calculating model option prices and fitting them to observed prices using Markov chain Monte Carlo would be too computationally intensive. Instead we follow the literature on model-free implied variance, on which indexes such as the VIX and the newly disseminated OVX are based, and calculate a non-parametric measure of implied variance from CME oil options. Define  $IV_{t,T}$  to be the implied variance of an asset between dates  $t$  and  $T > t$ . Specifically,  $IV_{t,T}$  is the risk-neutral expected value of the instantaneous return variance of the asset, averaged over the period between  $t$  and  $T$ . Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) demonstrate that, for a large class of diffusion processes, if the risk-free rate is not correlated with futures prices then<sup>5</sup>

$$(T - t) \times IV_{t,T} = E_t^* \left[ \int_t^T \left( \frac{dF_{s,T-s}}{F_{s,T-s}} \right)^2 \right] = 2 \int_0^\infty \frac{C_t(T - t, K) e^{y_{t,T-t}(T-t)} - (F_{t,T-t} - K)^+}{K^2} dK, \quad (3)$$

where  $C_t(\tau, K)$  is the date- $t$  price of a European call option with strike price  $K$  and maturity of  $\tau$  and  $y_{t,\tau}$  is the time- $t$  yield of a risk-free discount bond with maturity  $\tau$ . The integral above can be constructed directly from option prices, and thus the implied variance,  $IV_{t,T}$  is essentially observed (albeit, with error). In Appendix C it is shown that the affine structure of our model leads to a linear expression of implied variance in  $X_{v,t}$ :

$$IV_{t,T} = \gamma(t, T) + \delta(t, T)X_{v,t}. \quad (4)$$

Thus the information content about  $X_{v,t}$  in option prices can be extracted very efficiently.

**Equities** The value of a firm is the present value of all short-term and long-term projects. Thus, the stock prices of firms with high exposure to the oil risk factors should reflect information about both the long horizon and short horizon state variables. Unlike the case of futures and options, we do not have a structural model of how stock prices depend on oil prices. Instead, we rely on a simple model that is consistent with Merton’s (1973) intertemporal CAPM, the general equilibrium model of Cox, Ingersoll and Ross (1985), and other continuous-time versions of Ross’ (1976) APT where instantaneous stock returns are linear in the underlying state variables. To this end, recall that  $\Lambda_t$  from Eq. (1) is the instantaneous risk premium of  $X_t$ . The absence of arbitrage requires that a portfolio whose instantaneous returns are perfectly correlated with  $X_{i,t}$  must have instantaneous expected excess returns of  $\Lambda_{i,t}$ . If our model were perfectly specified *and* we observed futures and options prices without error, then we could construct such “factor-mimicking” portfolios and their instantaneous excess returns would be given by

$$dX_t^e = \left( dX_t - E_t[dX_t] \right) + \Lambda_t dt = \Sigma_t dW_t + \Lambda_t dt. \quad (5)$$

Our simple model for oil stock prices assumes that the returns on a portfolio of oil stocks will depend only on the market returns, on the oil factor-mimicking portfolios, and on non-systematic risk.<sup>6</sup> Specifically, let  $\frac{dP_t}{P_t} - r_{f,t}dt$  be the instantaneous returns of an oil stock portfolio in excess of the risk-free rate,  $r_{f,t}$ , and let  $\frac{dM_t}{M_t} - r_{f,t}dt$  be the instantaneous excess returns on the market portfolio. Then we model oil equity returns as,

$$\frac{dP_t}{P_t} - r_{f,t}dt = b_M \left( \frac{dM_t}{M_t} - r_{f,t}dt \right) + b'_{\text{oil}} dX_t^e + dz_t, \quad (6)$$

where  $dz_t$  is independent of  $dM_t$  and  $dX_t^e$ . The vector,  $b'_{\text{oil}} = (b_s, b_l, b_p, b_v)$  represents the oil factor sensitivities net of the effects of the market portfolio. In practice, one expects that the factor sensitivities  $b_M$  and  $b_{\text{oil}}$  will be time-varying. Absent a strong prior on the time-variation of these coefficients, we set them to be constants. This reduces our ability to extract information about the oil factors from stocks and constitutes a more conservative approach to the null hypothesis that stocks *do not* contain additional information about oil risk factors beyond the information contained in futures.

**Discretized Model** Data are observed only at discrete times and the transition density of the state vector must be computed over the length of time between observations. Because of the stochastic volatility, the transition density for our model is not known in closed form; however it can always be approximated by a Gaussian density by applying an Euler discretization. Although this approximation does introduce some error in the computed transition density, previous studies have shown the error is typically very small for observation frequencies less than roughly one week. We use daily observations in our empirical investigation and thus the error should be small. We also assume the observed futures and option prices are measured with white noise error,  $\xi_t$  and  $e_t$ , respectively. The complete discretized model that we estimate, written in state space form, is

$$X_{t+\Delta t} = A\Delta t + (I + B\Delta t)X_t + \sqrt{\Delta t}\Sigma_t \varepsilon_{t+\Delta t}, \quad (7)$$

$$\log F_{t,\tau} = \chi(\tau) + \sum_{i=s,l,p,v} \eta_i(\tau)X_{i,t} + \xi_{t,\tau}, \quad (8)$$

$$\text{IV}_{t,T} = \gamma(t, T) + \delta(t, T)X_{v,t} + e_{t,T}, \quad (9)$$

$$r_{z,t+\Delta t}^e = b_{z,M}R_{M,t+\Delta t}^e + \sum_{i=s,l,p,v} b_{z,i}R_{i,t+\Delta t}^e + \epsilon_{z,t+\Delta t}, \quad (10)$$

where  $\varepsilon_t, \xi_t, e_t, \epsilon_t$  are normal iid; the coefficient matrices are functions of the parameters and contract maturities;  $r_{z,t+\Delta t}^e$  and  $R_{M,t+\Delta t}^e$  denote excess returns on the equity portfolios indexed by  $z$  and on the market, respectively;  $R_{i,t+\Delta t}^e$  is the  $i$ th component of  $X_{t+\Delta t} - X_t - A^*\Delta t - B^*\Delta tX_t$ , which is the

discrete-time version of the excess-return on the factors in Eq. (5).

## II. Data and Empirical Methodology

The data consist of daily prices of crude oil (WTI) futures contracts as reported on the NYMEX, option contracts on WTI futures obtained from the CME, and daily returns of portfolios of oil-related firms. The sample period is 3/30/1983, when crude oil futures first began trading, to 12/31/2012. As discussed in Trolle and Schwartz (2009), the more liquid futures are the nearest six or seven contracts and those with delivery in December for up to about four years out. The last day to trade a particular contract is the third business day prior to the 25th (or the last business day preceding the 25th if the 25th is not a business day) of the month prior to the delivery month. Because the vast majority of contracts are closed out prior to delivery, the open interest declines sharply during the final two weeks of the contract's life. Acknowledging this, we discard the nearby contracts from the data sample, and use the next six monthly contracts and the next three December contracts in the estimation. The total number of daily observed futures prices in the sample is 59,836 and the maturities of the contracts used in the sample range from about nine weeks to four and a half years.<sup>7</sup> Daily volume and open interest for the shortest maturity series we use average 56,297 and 104,480 contracts, respectively.

Crude oil options on WTI futures began trading in November 1986 with each option expiring three business days prior to the last day to trade the underlying futures. Options trade far less actively than futures: In our sample options with maturities between 15 and 60 days averaged a daily volume and open interest of 614 and 3,985 contracts, respectively. We drop options with maturity of 14 or fewer days because their price data are less reliable (there is significantly less interest in the underlying contract). Using the remaining contracts, we construct the implied variance index in Eq. (3) with the nearest-to-maturity out-of-the-money contracts.<sup>8</sup> Details of the index construction are available in the Internet Appendix.

We use four equity portfolios in the estimation of Eqs. (7)-(10) constructed from the returns on

oil-related stocks. Specifically, each year we form quintiles of size-sorted stocks with SIC classification 1311, 1381, 1389, 2911, or 5172. We then value-weight all stocks in the top quintile to form one portfolio, and repeat for the bottom quintile. We similarly construct top and bottom quintile portfolios based on the book-to-market ratio.

To investigate the systematic nature of our factors we obtain from Kenneth French's online data library returns on 49 industry portfolios, the 25 Size and Book-to-market portfolios, the 25 Size and Momentum portfolios, the Fama-French-Carhart (see Fama and French, 1993, and Carhart, 1997) factors, and the risk-free rate.

We employ Markov chain Monte Carlo (MCMC) to estimate the model in Eqs. (7)-(10). Appendix D describes our procedure in greater detail. To assess the impact of using information from the equity market, we first estimate the model using only the futures and options data. We then repeat the estimation procedure including information from equity returns. For each estimation, a Markov chain of length 200,000 is generated, discarding the first 100,000 draws in the MCMC simulation to negate the effects of initial conditions. From the remaining draws, only one out of every ten is saved to improve the independence of each simulated instance, leaving 10,000 draws from the posterior distribution.

### **III. Results and Empirical Analysis**

Our analysis is divided into three parts. The first part is devoted to reporting key results from the estimation and goodness of fit. By comparing model estimates with and without equity returns, we establish that equity returns contain information about the oil risk factors that is not contained in oil futures and options. We find that, along with the market returns, the four oil factors explain a substantial amount of the variation in oil-related equity portfolios, especially when compared with traditional methods of accounting for systematic risk. We then establish that the model is robust by assessing out of sample goodness of fit. The second part considers the implications of the model for oil risk and return. Oil risk appears to be priced, and hence systematic. Moreover, the model is

instrumental in deriving estimates of oil risk premia, net of the market equity premium. The third part assesses the relevance of the model beyond oil derivatives and equities by examining the relationship of the imputed oil factors with macroeconomic variables and non-oil stocks. We find that oil risk is significantly related to current and future stock market volatility, changes in the rate of inflation, industrial production growth, changes in percentage unemployment, and GDP growth. This supports the view that extracted risk factors are systematic and have economy-wide influence. Furthermore, we find that the magnitude of exposure of non-oil stock portfolios to oil risk is, on average, a sixth that of the oil industry, with industries connected to commodities or manufacturing exhibiting the highest sensitivity.

#### A. Estimation

##### A.1. Goodness of Fit

Table I reports the fit to the data based on the parameter estimates and filtered state variables. We refer to an estimate of the parameters and latent variables as “derivatives-only” when the data consists only of the futures and implied variance series. We refer to the estimate as “full-information” when equity returns are also included in the estimation. The parameter estimates are reported in the Internet Appendix. For the futures, the pricing error is given by the difference between the fitted log-prices and actual log-prices. Following Trolle and Schwartz (2009), on each day we calculate the root mean squared error (RMSE) for the futures contracts in the estimation, which is the square-root of the average squared errors across all futures contracts on that day. The table reports the average of these daily RMSEs. We also calculate an unadjusted  $R^2$  for all 67,545 futures price predictions. Both derivatives-only and full information estimates deliver excellent fits to the futures data. In both cases the RMSE is roughly 0.28% and the  $R^2$  is essentially one. For comparison, Cortazar and Naranjo (2006) report a RMSE of 0.51% for their Gaussian three-factor model, Hughen (2010) reports a RMSE of 0.44% for his maximal  $A_1(3)$  model, and Trolle and Schwartz (2009) report a RMSE of 0.39% for

[Table I is about here]

the most general specification of their model; although these studies use different futures data than we do, the results suggest that misspecification error is not a major concern in our formulation. While the fit is excellent, the RMSE is still about ten times higher than recent bid-ask spreads on the most liquid contracts.

We also calculate  $R^2$  and RMSE for the implied variance. The RMSE is calculated as the square root of the time-series average of squared-errors and as such understates the goodness of fit relative to the method used by Trolle and Schwartz (2009).<sup>9</sup> While the fit is good (with an  $R^2$  of more than 98% with or without the equity data), it is not as precise as the fit to the futures. On the other hand, the RMSE is roughly of the order of magnitude of the bid-ask spread on the associated portfolio of options.<sup>10</sup> We can conclude that, given the limitations of the data and the model, the fit to both derivative prices is excellent.

The most important difference between estimating the model with versus without the equity data is seen in the fit to equity returns. The proportion of variance explained by the model for each of the four equity return series is documented in Table I for both estimation approaches as well as the FFC model. Using only derivative prices to estimate the parameters yields results that are on par with the standard FFC model with an  $R^2$  of about 50%. Indeed, much of the explanatory power comes in both instances from the market component,  $R_{M,t}^e$  which on average explains 40% of the variance. By contrast, the full-information estimation procedure yields a substantial improvement in explanatory power, averaging 75.3% of the variation across the four portfolios. This provides strong evidence that the model is picking up relevant risk factors. The improvement in  $R^2$  is largely driven by improved marginal  $R^2$ 's in the short-term factor  $X_{s,t}$  and the volatility factor  $X_{v,t}$  (and shared equally between them). We argue later that this may be because both depend on high-frequency information contained in equity returns.

Figure 1 plots a time series of the model-estimated factors using full-information estimates. The estimation produces a posterior distribution for the value of each factor on each date and the

corresponding figure depicts the mean of that distribution. Note that the plot for the volatility factor,  $X_{v,t}$  is hard to distinguish from observed implied variance over the sample period in which options are available to construct implied variance. Both the transient and strictly long-term oil risk components exhibit similar cyclical behavior. This is important and suggests that the model is capable of disentangling a single shock into distinct short and long-term components. The third, persistent, component clearly experienced a secular increase since roughly 2002. This too is intuitive and predicts that the high price of oil experienced over the past ten years is likely to remain high in the foreseeable future. Finally, the volatility component appears to experience acute periods of high volatility reflecting events causing great uncertainty in energy markets. Most notable among these are the collapse of the OPEC agreements and subsequent OPEC price wars in late 1985 and early 1986, the First Gulf War between late 1990 and early 1991, the Second Gulf War in March 2003, and the Global Financial Crisis of late 2008 until early 2009. Beyond that,  $X_{v,t}$  has undergone a high regime between 1999 and 2002 and has since stabilized, albeit at a higher level than in the 1990's. It is worth noting that the correlation between  $X_{v,t}$  and the GARCH(1,1) volatility estimate for next-to-nearby futures returns is 0.652 while the correlation between  $X_{v,t}$  and historical volatility is 0.696. Given that the correlation between the GARCH(1,1) volatility and historical volatility is a mere 0.463, we conclude that  $X_{v,t}$  adequately captures stochastic volatility in our estimation approach.

### *A.2. Model Stability*

To test the robustness of the estimation methodology, the sample period was divided into three parts: early (3/30/1983 to 12/31/1993), middle (1/3/1994 to 12/31/2003), and late (1/2/2004 to 12/31/2013). The top panel of Table II breaks down the fitting statistics by subperiod for the full-information estimates. The results suggest a high degree of fitting consistency for the pricing of derivatives across subsamples. The results for equity returns suggest that the assumption of constant factor loadings does not hold in practice.

We next estimated the model (using derivatives and equity returns) restricting the data to one of the three subsamples and, fixing the resulting parameters, estimated the latent variables in the other two subsamples. The lower panel in Table II reports the results in three subtables, corresponding to the subperiod used for the parameter estimates (within each subtable, “OOS” corresponds to an out of sample measure of fit). The results confirm the model’s stability for the pricing of derivatives. The fit to equity returns, on the other hand, appears to exhibit some deterioration in goodness of fit, also consistent with the conclusion that the factor loadings are not constant. Overall, estimating the model parameters using the late subsample appears to lead to the smallest amount of deterioration relative to in-sample estimates. Although the out of sample fit to equity returns declines when the early or middle period is used for estimating the parameters, we note that the resulting  $R^2$ ’s are nonetheless substantially better than those from Fama-French-Carhart regressions (the difference in  $R^2$  is consistently about 20%).

### *B. Oil Risk and Return*

Eq. (5) characterizes the instantaneous excess returns for a portfolio that is perfectly correlated with the oil factors. In discrete-time, we denoted these excess returns as  $R_{i,t}^e$  with  $i = s, l, p, v$ . As mentioned following Eq. (10),  $R_{i,t+\Delta t}^e$  is the  $i$ th component of  $X_{t+\Delta t} - X_t - A^*\Delta t - B^*\Delta t X_t$  and can therefore be explicitly calculated from our estimation of the model parameters and latent factors. The top panel of Table III reports on the time-series statistics of the daily  $R_{i,t}^e$ ’s estimated with full-information. Only the sample mean (i.e., sample risk premium) of  $R_{p,t}^e$ , is significantly different from zero. This is because, for the size of our sample period, a series of excess returns must have an annualized in-sample Sharpe Ratio of at least 36% to yield a significant sample mean of excess returns at the 5% confidence level. Only the persistent oil risk factor is associated with an in-sample Sharpe Ratio that is greater than 36%.<sup>11</sup> Thus, the relatively short time-series restricts one’s ability to directly measure small but economically meaningful levels of risk premia. Fortunately, we can calculate the

risk premia from the model parameters as follows. Eq. (1) identifies the conditional risk premium in terms of the estimated model parameters and latent factors. Correspondingly, the unconditional risk premium is given by the four-element vector

$$\bar{\Lambda} \equiv A - A^* + (B - B^*)\bar{X},$$

where  $\bar{X}$  is the unconditional mean of  $X_t$  (under the physical measure) and is a function of the estimated model parameters only. More importantly, the MCMC estimation delivers a posterior distribution for  $\bar{\Lambda}$ , thus allowing us to formally perform significance tests. This is akin to a Fama-MacBeth regression where risk premia are also estimated indirectly. In the latter case, sufficient cross-sectional variation in the expected returns of the basis assets is required to achieve statistical significance for risk premia associated with factors that do not have high in-sample Sharpe Ratios. In our case, the cross-sectional variation is provided by the different types of assets incorporated into the estimation methodology (i.e., futures, options, and stocks). The next-to-bottom row in the top panel of Table III documents the mean of the posterior distributions for the four elements of  $\bar{\Lambda}$ . Both the persistent and volatility risk factors are associated with economically *and* highly statistically significant risk premia. The corresponding model-implied unconditional Sharpe Ratios are documented in the last row and are likewise large for the persistent and volatility risk factors. The sample Sharpe Ratios are well within statistical tolerance of their model-implied values. It is important to mention that, when only derivative prices are used in the estimation, the model-implied risk premium and Sharpe Ratio for the volatility factor is not significantly different from zero, meaning that the inclusion of equities is critical in estimating risk premia more accurately. Overall, the evidence suggests oil risk is priced and therefore systematic.

The bottom panel of Table III reports the means of the posterior distribution for  $b_M$  and the  $b_i$ 's for  $i = s, l, p, v$ ; that is, the factor loadings of the return series on the market portfolio and on  $R_{i,t}^e$ 's (see Eq. (10)). The coefficients on the market and oil risk factors are highly significant and share similar

[Table III is about here]

signs across all portfolios. There appears to be little question that all of these factors are important in explaining the returns of the oil equity time series, although the cross-sectional variation across the loadings is modest. The contribution of the oil factors to the risk premia of the equity portfolios is roughly between 6%-7% per annum (calculated using the posterior distribution of  $\bar{\Lambda}$ ).

The positive sensitivity of oil stocks to the short-term, long-term, and persistent factors are intuitive. Less intuitive, however, is the sensitivity to the volatility factor. After controlling for the determinants of future oil value (i.e.,  $X_{s,t}$ ,  $X_{l,t}$  and  $X_{p,t}$ ) one might expect the real options possessed by oil-producing firms would lead to a *positive* loading on volatility, much in the same way that growth options might exhibit positive Vega. Instead, we find the opposite, suggesting that oil price volatility affects the value of oil-related firms in a non-trivial manner, or that our estimation methodology is forcing  $X_{v,t}$  to reflect something systematic other than oil volatility. As mentioned earlier, GARCH(1,1) and historical volatility estimates of the nearby log-futures price are less correlated with each other than they are with  $X_{v,t}$ . This suggests that  $X_{v,t}$  is an estimate of oil price volatility that is as good (or better) than either GARCH(1,1) or historical volatility. Moreover, the correlation with implied variance is extremely high and the RMSE is of the order of the bid-ask spread for an implied variance portfolio. Thus, if  $X_{v,t}$  is spuriously related to some systematic factor other than oil volatility, we are not able to identify what this systematic factor might be.<sup>12</sup> On the other hand, competition among international oil companies may be sufficient to diminish the positive effects of real options (e.g., Grenadier, 2002) leaving mostly the impact that volatility would have in the absence of real options.<sup>13</sup> In the next subsection we will establish that  $R_{v,t}^e$  is negatively related to indicators of economic growth and positively related to the VIX index. We speculate that this, together with competition in the oil industry, is sufficient to overwhelm any real-option effects. In other words, if a positive shock to  $X_{v,t}$  is sufficiently bad for the entire economy, it is also bad for the oil industry.

### *C. Oil Risk Beyond the Oil Industry*

The previous section demonstrates that the extracted oil factors explain oil-sensitive securities well. Such a link between the price of a product and stocks in its industry could be true within any industry, and perhaps is not surprising. Two of our factors,  $X_{p,t}$  and  $X_{v,t}$  carry economically significant risk premia and according to the APT must represent systematic risk. This, if true, must have ramifications outside the oil industry, which we investigate in this subsection. We do so, first, by showing that there is an empirically meaningful relationship between the oil volatility risk factor and macroeconomic variables. We then quantify the amount of systematic oil risk, and implied risk premium, to which non-oil portfolios are exposed (both industry portfolios and portfolios sorted on characteristics). Overall, we find non-oil companies are on average exposed to roughly a sixth of the magnitude of oil-specific risk to which the oil industry itself is exposed. Additionally, an analysis of oil factor loadings and oil risk premia for individual industry and oil-related sub-industry portfolios can be found in the Internet Appendix.

#### *C.1. Relationship to Macroeconomic Variables*

Panel A in Table IV reports contemporaneous correlations between various macroeconomic variables and the model-extracted oil factor excess returns between 1983 and 2012. At every date, the estimation procedure produces a distribution for the  $R_{i,t}^e$ 's and we use the mean of that distribution to calculate correlations or regression results.<sup>14</sup> We include the short rate (three-month constant-maturity Treasury rate), term premium (spread between Treasury 10-year and 3-month rates), default premium (spread between Moody's Baa and Aaa rates), change of squared VIX (squares and differences are taken for better comparison with  $R_{v,t}^e$ ), changes in inflation (growth rate of CPI), real growth of industrial production, changes in unemployment, and real GDP growth. All macro-economic variable data are annualized quarterly time series.<sup>15</sup> The daily factor excess returns are correspondingly compounded to produce quarterly time series. Of the four risk factors it is implied oil variance,  $R_{v,t}^e$ , that appears most

[Table IV is about here]

related to macroeconomic variables, exhibiting significant correlations with default spreads, inflation changes, changes in unemployment, and GDP growth. Other than inflation, the signs of these correlations make it clear that increases in  $X_{v,t}$  are linked to negative economic outcomes. The other oil factors are more weakly related to macroeconomic variables. In particular, while the third persistent factor is associated with a significant risk premium, it does not appear to be significantly related to any contemporaneous variable in the panel. On the other hand, one expects economic variables with oil sensitivity to be simultaneously related to a combination of the factors, which suggests that univariate correlations might not capture the extent of dependence on oil risk.

Pursuing this further, Panel B of Table IV documents the results of contemporaneous multivariate regressions of the macro variables on the oil risk factors. Several results stand out. First, the regressions confirm that key economic indicators have exposure to oil risk. Nearly half of the coefficients are significant at the 5% level or better. In particular, the volatility factor, which appears significant in five of the eight regressions, is unambiguously associated with bad news for the economy. This might be the reason that even oil companies' returns react negatively to increases in  $X_{v,t}$ : If competition in the oil industry dilutes the value of real options then the industry's procyclicality can result in a negative sensitivity to volatility. One can also see that the long-term and persistent factors are significantly linked to the macroeconomy, and especially market uncertainty (the VIX) and inflation. Finally, the varying sensitivities exhibited by the macro variables to the four oil factors justifies the need for decomposing oil shocks into its various components and suggests a path for clarifying ambiguous relationships found in previous studies (see Barsky and Kilian, 2004).

Turning to the predictive regressions (at 1 quarter lag) in Panel C, it appears that by and large only  $X_{v,t}$  is important in forecasting the various tested variables beyond the influence of the lagged macro variable itself.<sup>16</sup> That said, the economic significance of the predictability is impressive. A one (unconditional) standard deviation increase in  $R_{v,t}^e$  roughly predicts a 0.67% decline in annualized real GDP and a 1.7% decline in the growth of industrial production. Moreover, the 29% adjusted  $R^2$  in the

GDP predictive regression suggests that  $R_{v,t}^e$  compares well with other forecasters of real GDP growth at this horizon (e.g., see Ang, Piazzesi, and Wei, 2006). The highly significant predictive power of  $R_{v,t}^e$  for GDP growth, unemployment rate, and industrial production growth suggest that oil volatility shocks are not mere proxies for a business cycle factor.

### C.2. Oil Risk in Non-oil Stocks

**Fama-French-Carhart Factors** Panel A of Table V reports regressions of the oil factor excess returns on the Fama-French-Carhart 4-factor model. It is noteworthy that all four oil factors exhibit significant loadings, although only the short-term and volatility factors feature substantial  $R^2$ 's. Most impressive is the fact that the single highest loading is that of the volatility factor on the HML portfolio ( $-0.910$ , with a  $t$ -ratio of  $-22.6$ ), tying the oil volatility factor to the value premium.<sup>17</sup> It also appears that the risk premium for the volatility factor is explained by its relationship with the FFC factors. This contrasts with the persistent oil risk factor, which has only a weak relationship with the FFC factors but exhibits a significant daily alpha of  $0.055\%$  ( $t$ -stat of  $2.31$ ) amounting to about  $14\%$  per annum, leaving us without a clear explanation of the source of the risk premium for  $R_{p,t}^e$ .

Panel B of Table V documents coefficients from regressing market returns on the oil risk factors, as well as regressing the Fama-French oil industry returns and FFC factor returns on the oil factor and market returns. Oil risk is not strongly related to aggregate stock market risk, evidenced by the  $4.3\%$  adjusted  $R^2$  in the market regression. Each FFC model factor has significant though modest exposure to at least three oil risk factors (net of the market). The partial  $R^2$  statistic measures the variation explained by oil risk net of the market, and is calculated as the difference between the adjusted  $R^2$  using all five regressors and the adjusted  $R^2$  regressing only on market returns. Here we see that the oil factors capture a non-trivial amount of variation in HML and SMB (and more so than the market), potentially because of their link to business cycle indicators (Liew and Vassalou, 2000; and Petkova and Zhang, 2005).<sup>18</sup> The partial  $R^2$  for oil industry equity returns is comparable to that of the oil

[Table V is about here]

equity portfolios used in the estimation (see Table I).

The exposure of the FFC portfolios to oil risk factors can be measured relative to that of the oil industry. For  $i = s, l, p$  or  $v$ , and any asset excess return denoted by  $r_z$ , define

$$\beta_i(r_z) \equiv \frac{b_{z,i}}{|b_{\text{oil ind},i}|}, \quad (11)$$

where  $b_{z,i}$  is the regression coefficient obtained by regressing  $r_z$  on the market and oil factor returns as in Eq. (10), while  $b_{\text{oil ind},i}$  similarly corresponds to the regression coefficient for oil industry excess returns. In other words,  $\beta_i(r_z)$  measures the sensitivity of a portfolio's returns to oil risk relative to that of oil industry stocks. Columns 8-11 of Panel B, Table V report the oil betas for the SMB, HML, and UMD portfolios. Their magnitude averages to 0.17, corresponding to about a sixth of the exposure of the oil industry. Placing this into perspective, the weight of the oil industry portfolio in the market portfolio is roughly 3.1% throughout our sample period. This substantiates the notion that oil risk impacts the economy far beyond oil-related firms.

In the last column of Panel B, Table V we calculate the unconditional net oil risk premium for each of the portfolios. This is done using the posterior distribution of parameters and latent factors produced by the MCMC estimation. Specifically, we simulate the parameters and  $X_{i,t}$ 's from the posterior distribution, regress the FFC and oil industry returns on each corresponding instance of the  $R_{i,t}^e$ 's and the market excess return, randomly draw the normally distributed regression coefficients, and multiply the drawn oil factor regression coefficients by that simulation's  $\bar{\Lambda}$  to produce a possible realization of the risk premium that incorporates both model estimation *and* sample estimation errors. This generates a distribution of risk premia for each portfolio. The table reports the mean of the distribution and indicates whether it is significantly different from zero. About a sixth of the risk premia of SMB and HML appears to be captured by their oil risk exposures. This is consistent with an APT structure spanned by 5-7 fundamental risk factors of which oil is but one.

**Fama-French Portfolios** In Table VI we repeat the exercise of calculating oil risk betas (net of the market) and net oil premia for the 25 Size-B/M and the 25 Size-Momentum portfolios of Fama and French.<sup>19</sup> The oil betas documented in Panel B of Table V are reflected in the cross-sectional variation across the sorted portfolios. Smaller firms (value firms) tend to have a larger magnitude of exposure than larger firms (growth firms). Recent “winners ” have less negative exposure to oil volatility and more positive exposure to long-term oil risk than recent losers. All but three of the 50 portfolios have negative and significant exposure to the volatility factor. The average magnitude of exposure across all 50 portfolios is 20% that of the oil industry, confirming that oil risk plays a modest but economically significant role in the economy.

As in Table V, the net oil risk premium for each portfolio is calculated in the bottom panel (significant premia are in bold to make any pattern easier to identify). The average magnitude of net oil risk premium across the portfolios is 0.70% per annum which, while small, is consistent with the hypothesis that oil risk is one of a handful of key systematic risk factors in the economy. Note that it is typically larger-firm portfolios that feature significant risk premia despite generally having lower exposure to oil risk. This is because the larger-firm portfolios are less volatile and their factor loadings can be determined more accurately.

The greatest cross-sectional variation in oil factor loadings is between Size=1 & B/M=5 portfolio and the Size=5 & B/M=1 portfolio — among the 25, it is also these two portfolios that exhibit the highest cross-sectional dispersion in expected return. A portfolio long on the former and short the latter has an average magnitude of oil betas that is about one half of that of the oil industry and a net oil risk premium of 1.90%. The Size=5 & B/M=1 portfolio is particularly interesting in that it appears to act as a hedge against oil risk.

**Industry Portfolios** The preceding analysis confirms that our extracted oil factors are systematic and that portfolios formed on financial characteristics vary cross-sectionally with respect to oil risk.

In particular, firms whose financial characteristics make them particularly sensitive to business cycle variables appear to have greater net exposure to oil risk. Next we examine a different cross-sectional property: industry. The FFC model is not generally as useful when applied to industry portfolios, possibly because industry portfolio returns do not vary enough across size and B/M. One does intuitively expect, however, that some industries will have significantly higher sensitivity to oil risk than others. To pursue this intuition, we use the 49 industry returns collected by Kenneth French and posted on his website, and remove the oil industry portfolio used earlier in Panel B of Table V. Each day and for each industry, starting in 6/24/1983 and ending in 12/31/2012, we regress the past 60 trading days of industry excess returns on the market excess return and the  $R_{i,t}^e$ 's (also from the past 60 trading days). On a daily basis, we then sort the 48 industry portfolios based on the regression coefficient of  $R_i^e$  (we do this separately for  $i = s, l, p$  and  $v$ ). Next, we form 8 “octile” portfolios out of that day’s loading-sorted industry returns (e.g., the lowest octile equally-weights the returns of the 6 industries with the lowest regression coefficient). Altogether, we construct  $8 \times 4$  portfolios: eight octiles rebalanced daily and based on the regression coefficient with respect to each of the four factors.

If there was no cross-sectional variation in net oil risk exposure across industries, then each industry would be represented equally in each octile and appear with frequency 2.08%. Table VII documents the five highest-frequency industries in the first and last octiles across all days. A clear pattern emerges: precious metals, mining, and coal consistently appear in octile 8 for  $R_s^e, R_l^e$  and  $R_p^e$ , and correspondingly in octile 1 for  $R_v^e$  (recall that exposure to  $R_v^e$  is generally negative for firms that are positively exposed to the other oil factors). Likewise, construction appears three times and steel twice in these same octiles. Thus, the industries with highest oil sensitivity are commodities and basic manufacturing. The least oil-sensitive industries appearing consistently (in octile 1 for  $R_s^e, R_l^e$  and  $R_p^e$ , and in octile 8 for  $R_v^e$ ) are tobacco products, candy & soda, computer hardware, defense, and electronic equipment.

[Table VII is about here]

Table VIII reports the net oil betas and risk premia of the 32 sensitivity-sorted portfolios. The

table confirms that sorting based on the sensitivity of one factor is highly correlated with sorting based on another factor (negatively correlated, when sorting on  $R_v^e$  sensitivity). In other words, most firms that are highly sensitive to one risk factor tend to also be proportionately sensitive to the others. Moreover, only the top two octiles (bottom two octiles when sorting on  $R_v^e$  sensitivity) admit a high magnitude of sensitivity to the oil factors. Thus, in practice, about 25% of industries are substantially affected by oil risk, although which industries are affected can vary over time. In both Tables VI and VIII, the preponderance of positive betas with respect to the first three factors and negative betas for the volatility factor suggest that relatively few stocks can be used to hedge the oil risk factors. Thus, because oil risk is hard to hedge or completely diversify away, it should be priced. Indeed, the average magnitude of the betas is 14.3% while the average magnitude of risk premium is 0.70%. Averaging the magnitude of exposures of the 32 industry octile portfolios and the 50 Fama-French portfolios from Table VI yields an average magnitude of exposure of roughly one sixth that of the oil industry. Finally, despite the fact that only non-oil industries are used to construct the octile portfolios, a portfolio long octile 8 and short octile 1 (the “8 minus 1” column) will have oil betas that are roughly two thirds in magnitude relative to that of the oil industry and a commensurately high net oil risk premium. Moreover, while each of the “8 minus 1” portfolio risk premia are significant at the 1% level, the hypothesis that all four have zero net-oil risk premium can be rejected at the 0.1% level (accounting for their correlations), consistent with the more demanding threshold in Harvey, Liu and Zhu (2013).

In the Internet Appendix we report individual industry and oil sub-industry loadings and risk premia. Sixteen out of 49 exhibit significant risk premia with commodity-linked industries having the largest risk positive premia while retail, technology, and pharmaceuticals having the most negative. Among oil sub-industries, airlines have a large and negative exposure with a net oil risk premium of  $-4.88\%$  coming mostly from the persistent factor. Oil equipment and services is the most sensitive sub-industry with 1.80 times the volatility exposure of the oil industry and a net oil risk premium of  $8.24\%$ .

[Table VIII is about here]

### *C.3. Oil Risk Information in Oil Stocks*

Do the factors explain as much of the cross-section when they are estimated using derivative prices only (“DO oil factors”)? When applying the analysis in this section to DO factors we find that *none* of the 139 portfolios examined in Tables V , VI and VIII exhibit risk premia that are significant at the 1% level (8 are significant at the 5% level). This is despite the fact that the factors do exhibit strong evidence of priced (and therefore systematic) risk. Thus, the equities used in the estimation are crucial in capturing the covariation of stocks with oil risk. By contrast, Table IV when constructed using the DO oil factors produces results that are qualitatively similar to what is obtained with the full information (FI) estimation of the factors. Indeed, what seems to be going on is that the DO and FI oil factors estimates are poorly correlated at daily frequencies but very highly correlated at quarterly frequencies. Thus equities contain high frequency information about the factors. This can be because daily derivative price fluctuations are noisy — whether because of modeling error (we are not incorporating embedded timing options) or because of short-term market segmentation (the two markets are linked but short-term deviations between them is hard to arbitrage away).

## **IV. Conclusion**

We study the implications of commodity derivative models for equity markets. We do this by including oil equity returns in the estimation of a structural model of the term structure of oil futures with unspanned volatility. The estimation technique delivers a time series for four oil risk factors, corresponding to shocks influencing oil prices in the short-term, in the long-term, persistently, or through volatility shocks. Our fit to derivative prices is excellent and robust, while the risk factors we identify appear to be associated with a significant risk premium. When information from oil-sensitive equities is incorporated into the estimation of the oil factors, their ability to explain oil-related stock returns drastically improves (the  $R^2$  increases by an average of roughly 20%) while the fit to derivatives prices remains largely unchanged. Moreover, the additional information from equities is crucial in

allowing us to better estimate the oil factor risk premia.

The extracted oil risk factors, and in particular the one corresponding to unspanned volatility, appear to be systematic. Oil volatility exhibits an economically and statistically significant negative relationship with growth variables such as real GDP, industrial production and a positive relationship with negative indicators such as the VIX Index and unemployment. The magnitude of oil risk exposure for an average non-oil firm is between 14% and 20% of the oil risk exposure for the oil industry itself, and corresponds to an average risk premium magnitude of 0.70%.

Overall, our results provide evidence that oil risk (or alternatively, energy risk), adequately filtered from noisy price data, is one of a handful of fundamental factors that affect the pricing of both oil and non-oil securities. In principle, our methodology could be applied to extracting other fundamental risk factors, paving the way for a more structured approach to the pricing of securities in an APT framework.<sup>20</sup>

## References

- Al-Mudhaf, Anwar, and Thomas H. Goodwin, 1993, Oil shocks and oil stocks: Evidence from the 1970s, *Applied Economics* 25, 181-190.
- Ang, Andrew, Monika Piazzesi, and Min Wei, 2006, What does the yield curve tell us about GDP growth? *Journal of Econometrics* 131, 359-403.
- Backus, David K., and Mario J. Crucini, 2000, Oil prices and the terms of trade, *Journal of International Economics* 50, 185-213.
- Baddour, J.W., 1997, The international petroleum industry: Competition, structural change and allocation of oil surplus, *Energy Policy* 25, 143-157.
- Baker, Steven D., and Bryan R. Routledge, 2012, The price of oil risk, Working paper, Carnegie Mellon University.
- Barsky, Robert, and Lutz Kilian, 2004, Oil and the macroeconomy since the 1970s, *Journal of Economic Perspectives* 18, 115-134.
- Britten-Jones, Mark, and Anthony Neuberger, 2000, Option prices, implied price processes, and stochastic volatility, *Journal of Finance* 55, 839-866.
- Campbell, John Y., Karine Serfaty-de Medeiros, and Luis M. Viceira, 2010, Global currency hedging, *Journal of Finance* 65, 87-121.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
- Casassus, Jaime, and Pierre Collin-Dufresne, 2005, Stochastic convenience yield implied from commodity futures and interest rates, *Journal of Finance* 60, 2283-2331.
- Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383-403.

- Cheridito, Patrick, Damir Filipovic, and Robert L. Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123-170.
- Collin-Dufresne, Pierre, Robert S. Goldstein, and Christopher S. Jones, 2009, Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* 94, 47-66.
- Cortazar, Gonzalo, and Lorenzo Naranjo, 2006, An  $N$ -factor Gaussian model of oil futures prices, *Journal of Futures Markets* 26, 243-268.
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-384.
- de Jong, Piet, and Neil Shephard, 1995, The simulation smoother for time series models, *Biometrika* 82, 339-350.
- Dumas, Bernard, and Bruno Solnik, 1995, The world price of foreign exchange risk, *Journal of Finance* 50, 445-479.
- Elyasiani, Elyas, Iqbal Mansur, and Babatunde Odusami, 2011, Oil price shocks and industry stock returns, *Energy Economics* 33, 966-974.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.
- Ferson, Wayne E., and Campbell R. Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527-566.
- Gibson, Rajna, and Eduardo S. Schwartz, 1990, Stochastic convenience yield and the pricing of oil contingent claims, *Journal of Finance* 45, 959-976.

- Grenadier, Steven R., 2002, Option exercise games: An application to the equilibrium investment strategies of firms, *Review of Financial Studies* 15, 691-721
- Hamilton, James D., 2003, What is an oil shock? *Journal of Econometrics* 113, 363-398.
- Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2013, ...And the cross-section of expected returns, Working paper, Duke University.
- Huang, Roger D., Ronald W. Masulis, and Hans R. Stoll, 1996, Energy shocks and financial markets, *Journal of Futures Markets* 16, 1-27.
- Hughen, W. Keener, 2010, A maximal affine stochastic volatility model of oil prices, *Journal of Futures Markets* 30, 101-133.
- Jiang, George J., and Yisong S. Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305-1342.
- Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics* 57, 221-245.
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.
- Oberndorfer, Ulrich, 2009, Energy prices, volatility, and the stock market: Evidence from the Eurozone, *Energy Policy* 37, 5787-5795.
- Petkova, Ralitsa, and Lu Zhang, 2005, Is value riskier than growth? *Journal of Financial Economics* 78, 187-202.
- Ramos, Sofia B., and Helena Veiga, 2011, Risk factors in oil and gas industry returns: International evidence, *Energy Economics* 33, 525-542.
- Ready, Robert C., 2012, Oil Prices and long-run risk, Working paper, University of Rochester.

Ready, Robert C., 2013, Oil Prices and the stock market, Working paper, University of Rochester.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

Sadorsky, Perry, 2001, Risk factors in stock returns of Canadian oil and gas companies, *Energy Economics* 23, 17-28.

Scholtens, Bert, and Lei Wang, 2008, Oil risk in oil stocks, *Energy Journal* 29, 89-111.

Schwartz, Eduardo, 1997, The stochastic behavior of commodity prices: Implications for valuations and hedging, *Journal of Finance* 52, 922-973.

Schwartz, Eduardo, and James E. Smith, 2000, Short-term variations and long-term dynamics in commodity prices, *Management Science* 46, 894-911.

Trolle, Anders, and Eduardo Schwartz, 2009, Unspanned stochastic volatility and the pricing of commodity derivatives, *Review of Financial Studies* 22, 4423-4461.

# Appendix

## Appendix A. The General $A_1(4)$ Model

The most general  $A_1(4)$  model for the risk-neutral evolution of the spot price process  $S$  is

$$\log S_t = a + b'Y_t, \tag{A1}$$

where  $Y_t \in \mathbb{R}^4$  has affine drift and instantaneous variance:

$$dY_t = (A^* + B^*Y_t)dt + \Sigma_t dW_t^*, \tag{A2}$$

$$\Sigma_t \Sigma_t' = \Omega_0 + \Omega_1 \times c'Y_t. \tag{A3}$$

The total number of parameters is 48: 1 in  $a$ , 4 in  $b$ , 4 in  $A^*$ , 16 in  $B^*$ , 10 in  $\Omega_0$ , 10 in  $\Omega_1$ , and because  $c$  is defined only up to a scalar multiple, 3 in  $c$ . However, these parameters aren't all well-defined because there is ambiguity in the choice of state variable  $Y$ : any invertible affine transformation  $X_t = \psi + \Phi Y_t$ , where  $\psi$  is a 4-vector and  $\Phi$  is an invertible  $4 \times 4$  matrix, will also have affine dynamics Eqs. (A1)–(A3). The group of symmetries of the set of state variables with  $A_1(4)$  dynamics is therefore 20-dimensional; intuitively, this represents 20 degrees of ambiguity in the choice of state variable  $Y$ , or equivalently, 20 degrees of ambiguity in the choice of parameter set  $\Theta = (a, b, A^*, B^*, \Omega_0, \Omega_1, c)$ , and therefore the identifiable model should have at most  $28 = 48 - 20$  free parameters.

Without loss of generality it may always be assumed that  $Y_{1,t} = \log S_t$  and  $Y_{4,t} = V_t$ , the (squared) volatility of  $S_t$ , by first replacing  $a + bY_t$  with  $Y_{1,t}$  and then replacing  $(\Omega_0)_{1,1} + (\Omega_1)_{1,1}c'Y_t$  with  $Y_{4,t}$ . After this substitution is made, the risk-neutral drift of  $Y_{1,t}$  may be replaced by  $Y_{2,t} - \frac{1}{2}Y_{4,t}$ , and then the risk-neutral drift of  $Y_{2,t}$  may be replaced by  $Y_{3,t}$ . With this identification, there is no more ambiguity in the choice of state variable; the group of symmetries now consists of only the trivial

transformation  $\psi = 0$  and  $\Phi = I$ , the identity matrix.

Therefore, given the most general affine model satisfying equations Eqs. (A1)–(A3) there exists a unique invertible affine transformation of the state vector for which  $Y_{1,t}$  is the log spot price,  $Y_{4,t}$  is the spot price volatility,  $\Sigma_t \Sigma_t' = \Omega_0 + \Omega_1 Y_{4,t}$ , and

$$A^* = \begin{bmatrix} 0 \\ 0 \\ u_1 \\ u_2 \end{bmatrix}, \quad B^* = \begin{bmatrix} 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix}.$$

A quick count reveals there are indeed 28 free parameters (2 in  $A^*$ , 8 in  $B^*$ , and 9 in each of  $\Omega_0$  and  $\Omega_1$  since  $\Omega_0(1,1) = 0$  and  $\Omega_1(1,1) = 1$ ), but admissibility will impose restrictions on some of these parameters. In particular, the drift of the volatility  $Y_{4,t}$  must be non-negative for the volatility to remain positive, and therefore  $k_{41} = k_{42} = k_{43} = 0$ . In addition, when the volatility state variable  $Y_{4,t}$  approaches zero, its instantaneous variance and covariance with the other state variables must vanish otherwise the volatility would drop below zero with positive probability, and therefore the fourth row and column of  $\Omega_0$  are zero.

For purposes of this study, it is more convenient to consider the equivalent representation

$$\log S_t = X_{1,t} + X_{3,t}$$

$$dX_t = (A^* + B^* X_t)dt + \Sigma_t dW_t^*$$

$$A^* = \begin{bmatrix} 0 \\ \mu_1 \\ 0 \\ \mu_2 \end{bmatrix}, \quad B^* = \begin{bmatrix} -\kappa_1 & 0 & 0 & \kappa_5 \\ 0 & -\kappa_2 & \kappa_4 & \kappa_1 \kappa_5 \\ 0 & 1 & 0 & -\frac{1}{2} - \kappa_5 \\ 0 & 0 & 0 & -\kappa_3 \end{bmatrix}$$

$$\Sigma_t \Sigma_t' = \Omega_0 + \Omega_1 X_{4,t}$$

where  $X_{4,t}$  is again the spot price volatility, so that

$$\Omega_0(1, 1) + 2\Omega_0(1, 3) + \Omega_0(3, 3) = 0$$

$$\Omega_1(1, 1) + 2\Omega_1(1, 3) + \Omega_1(3, 3) = 1,$$

and the fourth row and column of  $\Omega_0$  are zero. A straightforward calculation shows that given such a state variable  $X$ , the state variable given by

$$Y_t = \begin{bmatrix} 0 \\ 0 \\ \mu_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 0 \\ -\kappa_1 & 1 & 0 & 0 \\ \kappa_1^2 & -\kappa_2 & \kappa_4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} X_t$$

has dynamics congruent with the previous ‘ $Y$ -representation’ and therefore these two representations are equivalent. Further details are given in the Internet Appendix. We now show how USV and the presence of a persistent component further constrains the representation.

## Appendix B. Futures Prices and USV

Standard arguments from affine term structure models used to price bonds imply that the log futures price at time  $t$  for delivery at time  $t + \tau$  is given by

$$\log F_{t,\tau} = \chi(\tau) + \eta'(\tau)X_t,$$

where  $\chi$  and  $\eta$  satisfy the set of differential equations (with respect to  $\tau$ )

$$\dot{\chi} = \eta' A^* + \frac{1}{2} \eta' \Omega_0 \eta \quad (\text{B1})$$

$$\dot{\eta} = \eta' B^* + (0, 0, 0, \frac{1}{2} \eta' \Omega_1 \eta) \quad (\text{B2})$$

with initial values  $\chi(0) = 0$  and  $\eta(0)' = (1, 0, 1, 0)$ .

The model will exhibit USV if and only if the fourth component  $\eta_4$  is identically zero which from Eq. (B2) is equivalent to the condition

$$(\eta_1, \eta_2, \eta_3) \cdot (\kappa_5, \kappa_1 \kappa_5, -\frac{1}{2} - \kappa_5) + \frac{1}{2} (\eta_1, \eta_2, \eta_3, 0) \Omega_1 (\eta_1, \eta_2, \eta_3, 0)' \equiv 0. \quad (\text{B3})$$

In concordance with the premise of this study, we impose the restriction  $\kappa_4 = 0$  so that the log spot price has a persistent component (given by  $X_3$ ).<sup>21</sup> It follows that the first three components of  $\eta$  are given by

$$\eta_1(\tau) = e^{-\kappa_1 \tau}, \quad \eta_2(\tau) = \frac{1 - e^{-\kappa_2 \tau}}{\kappa_2}, \quad \eta_3(\tau) = 1;$$

substituting into Eq. (B3) and collecting terms reveals

$$\Omega_1(1, 1) = \Omega_1(1, 2) = \Omega_1(2, 2) = 0, \quad \kappa_1 \kappa_5 + \Omega_1(2, 3) = 0, \quad \kappa_5 + \Omega_1(1, 3) = 0.$$

To be well defined, the instantaneous variance  $\Omega_0 + \Omega_1 X_{4,t}$  must be positive definite for all  $X_{4,t} > 0$ ; it suffices that  $\Omega_0$  and  $\Omega_1$  are positive semidefinite and have disjoint null spaces. Because the upper left  $2 \times 2$  submatrix of  $\Omega_1$  is zero,  $\Omega_1$  will be positive semidefinite if only if  $\Omega_1(1, 3) = \Omega_1(2, 3) =$

$\Omega_1(1, 4) = \Omega_1(2, 4) = 0$  also, in which case  $\kappa_5 = 0$  as well. Thus

$$B^* = \begin{bmatrix} -\kappa_1 & 0 & 0 & 0 \\ 0 & -\kappa_2 & 0 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & -\kappa_3 \end{bmatrix}, \quad \Omega_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \sigma_{14} \\ 0 & 0 & \sigma_{14} & \sigma_{44} \end{bmatrix}, \quad \Omega_0 = \begin{bmatrix} s_{11} & s_{12} & s_{13} & 0 \\ s_{12} & s_{22} & s_{23} & 0 \\ s_{13} & s_{23} & s_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where  $s_{11} + 2s_{13} + s_{33} = 0$ . This restriction implies that the characteristic polynomial of  $\Omega_0$  has no negative root, that is,  $\Omega_0$  is positive semidefinite, if and only if

$$\Omega_0 = \begin{bmatrix} s_{11} & s_{12} & -s_{11} & 0 \\ s_{12} & s_{22} & -s_{12} & 0 \\ -s_{11} & -s_{12} & s_{11} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Furthermore,  $\Omega_0$  and  $\Omega_1$  will have disjoint null spaces if and only if the upper left  $2 \times 2$  submatrix of  $\Omega_0$  is strictly positive definite. See the Internet Appendix for more details.

Therefore, the risk-neutral model, with the restriction that  $\kappa_4 = 0$ , will be identifiable and admissible and will exhibit USV if and only if

$$dX_t = \left( \begin{bmatrix} 0 \\ \mu_1 \\ 0 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} -\kappa_1 & 0 & 0 & 0 \\ 0 & -\kappa_2 & 0 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & -\kappa_3 \end{bmatrix} X_t \right) dt + \Sigma_t dW_t^*$$

$$\Sigma_t \Sigma_t' = \Omega_0 + \Omega_1 X_{4,t}$$

$$\Omega_0 = \begin{bmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & -\sigma_1^2 & 0 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & -\rho_1\sigma_1\sigma_2 & 0 \\ -\sigma_1^2 & -\rho_1\sigma_1\sigma_2 & \sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \Omega_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho_2\sigma_3 \\ 0 & 0 & \rho_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

where  $-1 < \rho_1, \rho_2 < 1$ ,  $\kappa_1, \kappa_2, \kappa_3, \sigma_1, \sigma_2, \sigma_3 > 0$  and  $2\mu_2 > \sigma_3^2$ . Trivial relabeling of the parameters leads to the set of expressions for the model and futures prices in Section I.

### Appendix C. Model-free Implied Variance

The implied variance

$$\text{IV}_{t,T} = \frac{1}{T-t} E_t^* \left[ \int_t^T \left( \frac{dF_s}{F_s} \right)^2 \right]$$

is constructed directly from futures option prices following Britten-Jones and Neuberger (2000) and Jiang and Tian (2005), where  $T$  is the maturity date of the option and the underlying futures contract matures at time  $T^* > T$ . Due to the affine structure, the volatility of futures prices is

$$\left( \frac{dF_s}{F_s} \right)^2 = \eta(T^* - s)(\Omega_0 + \Omega_1 X_{v,s})\eta(T^* - s)' ds$$

and therefore

$$(T-t) \times \text{IV}_{t,T} = \int_t^T \eta(T^* - s)\Omega_0\eta(T^* - s)' ds + \int_t^T \eta(T^* - s)\Omega_1\eta(T^* - s)' E_t^*[X_{v,s}] ds.$$

The first integral is straightforward. The second integral can be computed by noting that  $\eta(T^* - s)\Omega_1\eta(T^* - s)' = 1$  and  $E_t^*[X_{v,s}] = e^{-\kappa_v(s-t)}X_{v,t} + \frac{\mu_v}{\kappa_v}(1 - e^{-\kappa_v(s-t)})$ . It follows that the implied variance is of the form

$$\text{IV}_{t,T} = \gamma(t, T) + \delta(t, T)X_{v,t} \tag{C1}$$

where  $\gamma$  and  $\delta$  can be found by computing the required integrals:

$$\begin{aligned}\gamma(t, T) &= \frac{1}{T-t} \int_t^T \eta(T^* - s) \Omega_0 \eta(T^* - s)' + \frac{\mu_v}{\kappa_v} (1 - e^{-\kappa_v(s-t)}) ds \\ \delta(t, T) &= \frac{1}{\kappa_v(T-t)} (1 - e^{-\kappa_v(T-t)}).\end{aligned}$$

#### Appendix D. MCMC Estimation

Under our assumptions of admissibility, stationary risk premium, and USV, the state space form of the model is

$$\begin{aligned}X_{t+\Delta t} &= A\Delta t + (I + B\Delta t)X_t + \sqrt{\Delta t}\Sigma_t \varepsilon_{t+\Delta t} \\ \log F_{t,\tau} &= \chi(\tau) + \eta(\tau)'X_t + \xi_{t,\tau} \\ \text{IV}_{t,T} &= \gamma(t, T) + \delta(t, T)X_{v,t} + e_{t,T} \\ r_{z,t+\Delta t}^e &= b_{z,M}R_{M,t+\Delta t}^e + b'_{z,\text{oil}}(X_{t+\Delta t} - X_t - A^*\Delta t - B^*\Delta t X_t) + \epsilon_{z,t+\Delta t}.\end{aligned}$$

Because the returns each period are linear functions of the current and lagged state variables, it is convenient to consider the stacked state vector  $\tilde{X}_t = \begin{bmatrix} X_t \\ X_{t-\Delta t} \end{bmatrix}$ , so that the transition and observation equations become

$$\tilde{X}_{t+\Delta t} = \begin{bmatrix} A\Delta t \\ 0 \end{bmatrix} + \begin{bmatrix} I + B\Delta t & 0 \\ I & 0 \end{bmatrix} \tilde{X}_t + \sqrt{\Delta t} \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix} \tilde{\varepsilon}_{t+\Delta t} \quad (\text{D1})$$

and

$$Z_t = \begin{bmatrix} \chi_t \\ \gamma_t \\ b_M R_{M,t}^e - b'_{\text{oil}} A^* \Delta t \end{bmatrix} + \begin{bmatrix} \eta'_t & 0 \\ \tilde{\delta}_t & 0 \\ b'_{\text{oil}} & -b'_{\text{oil}}(I + B^* \Delta t) \end{bmatrix} \tilde{X}_t + \tilde{\xi}_t \quad (\text{D2})$$

where  $Z_t$  denotes the vector of stacked log-futures prices, implied variances, and portfolio returns at time  $t$ ,  $\chi_t$  and  $\gamma_t$  are the stacked  $\chi(\tau)$  and  $\gamma(t, T)$  corresponding to the various futures and implied variance maturities used in the estimation,  $\eta_t$  is the array of stacked  $\eta(\tau)$  for the various futures maturities, and  $\tilde{\delta}_t$  is the 4-column array whose first three columns are zero and whose fourth column is the stacked  $\delta(t, T)$  for the various implied variance maturities. The stacked state vector is not Gaussian because the volatility state variable  $X_{v,t}$  appears in the conditional variance matrix; however, as noted in Hughen (2010) and Collin-Dufresne, Goldstein, and Jones (2009), conditional on  $X_v$  the remaining components of the (stacked) state vector are Gaussian with time-varying but deterministic variance. Therefore these remaining components may be drawn in a single block using, for example, the simulation smoother of de Jong and Shephard (1995).

The volatility state variable  $X_{v,t}$  is drawn using a separate Metropolis-Hastings step. By the law of total probability, the conditional density for  $X_{v,t}$  given all other state variables (including the volatility state variable at all other times), the futures, implied variance, and returns data, and the parameters is proportional to

$$p(\text{IV}_t | X_{v,t}, \Theta) p(r_t^e | X_t, X_{t-\Delta t}, \Theta) p(r_{t+\Delta t}^e | X_t, X_{t+\Delta t}, \Theta) p(X_{t+\Delta t} | X_t, \Theta) p(X_t | X_{t-\Delta t}, \Theta).$$

The first kernel  $p(\text{IV}_t | X_{v,t}, \Theta)$  is the (Gaussian) likelihood of the implied variance data viewed as a function of  $X_{v,t}$ , and the second and third kernels  $p(r_t^e | X_t, X_{t-\Delta t}, \Theta)$  and  $p(r_{t+\Delta t}^e | X_t, X_{t+\Delta t}, \Theta)$  are the (Gaussian) likelihoods of the return data on date  $t$  and  $t + \Delta t$ , viewed as functions of  $X_{v,t}$ , with all other parameters, the values of the other state variables, and the volatility state variable at all other times treated as fixed. The fourth and fifth kernels are the transition densities over the current and previous periods, respectively. As a function of  $X_{v,t}$  the conditional density is not a recognizable distribution; the volatility state variable  $X_{v,t}$  is drawn using independence Metropolis-Hastings with the Gaussian proposal density  $p(\text{IV}_t | X_{v,t}, \Theta) p(r_t^e | X_t, X_{t-\Delta t}, \Theta) p(r_{t+\Delta t}^e | X_t, X_{t+\Delta t}, \Theta) p(X_t | X_{t-\Delta t}, \Theta)$ . If

a negative candidate for  $X_{v,t}$  is drawn, it is simply discarded and a new candidate is drawn.

The parameter vector  $\Theta$  is decomposed into five blocks:  $\Theta_1 = (A, B)$ ,  $\Theta_2 = (\mu_l, \mu_v)$ ,  $\Theta_3 = (\kappa_s, \kappa_l, \kappa_v, \sigma_s, \sigma_l, \sigma_v, \rho_{sl}, \rho_{vp})$ ,  $\Theta_4 = (b_M, b_{oil})$ , and  $\Theta_5 = W$  is the variance matrix of the pricing errors. The first block  $\Theta_1$  appears only in the drift term of the state vector and appears linearly; therefore its conditional density is Gaussian and may be drawn in one step from a Gaussian distribution using the Gibbs sampler. Similarly, because the second block  $\Theta_2$  appears linearly in  $\chi_t, \gamma_t$ , and the risk-neutral drift terms  $A^*$  and  $B^*$ , it appears linearly in the observation equation Eq. (D2) and thus it may also be drawn in a single step directly from a Gaussian distribution. Likewise, the factor loadings  $b_M$  and  $b_{oil}$  appear linearly in Eq. (D2) and thus the fourth block  $\Theta_4$  may also be drawn in a single step directly from a Gaussian distribution. The fifth block  $\Theta_5$  may also be drawn in a single step. We assume homoscedastic variance of the futures pricing error, the implied variance error, and the variance of the idiosyncratic return error. By a standard result the likelihood function of  $W$  is an Inverse Gamma density function and therefore this block may be drawn directly using Gibbs sampling.

The third block is more difficult to draw because these parameters appear nonlinearly in Eq. (D2). As a result, the conditional density of  $\Theta_3$  is not a recognizable density function. This block is drawn using the relatively inefficient random walk Metropolis-Hastings with a Gaussian proposal density.

## Footnotes

<sup>1</sup> Attempts to use Fama and MacBeth (1973) type regressions to deduce oil risk premia from the cross-section of asset prices have yielded insignificant results (e.g., Chen, Roll, and Ross, 1986). Direct estimates of risk premia for oil factors using oil derivative price dynamics have yielded mixed results, but even when they have come up significant there has been no attempt to connect such results to cross-sectional asset pricing. Confounding this, Huang, Masulis, and Stoll (1996) find virtually no relation between futures returns and U.S. stock market returns. On the other hand, Ferson and Harvey (1993) find that a five-factor model including oil risk performs better than a single factor model in predicting the returns of global equity markets. Complementing this, Backus and Crucini (2000) observe that oil shocks are important determinants of relative price movements in international markets.

<sup>2</sup>Related derivative-pricing models include Casassus and Collin-Dufresne (2005), Gibson and Schwartz (1990), Hughen (2010), Schwartz (1997), Schwartz and Smith (2000), and Trolle and Schwartz (2009). Recent studies linking oil prices and price volatility to oil equities include Elyasiani, Mansur, and Odusami (2011), Oberndorfer (2009), and Ramos and Veiga (2011) — our more structural approach to extracting the factors allows us to additionally: obtain a better fit to oil equity prices over a longer horizon, simultaneously fit to oil derivative prices, uncover the risk premia associated with the oil factors, link oil factors to the macroeconomy, and demonstrate the factors' relationship with the broader cross section of equities.

<sup>3</sup>Similar assumptions about the relation between the risk-adjusted and true drifts are made in Cheridito, Filipovic, and Kimmel (2007), Collin-Dufresne, Goldstein, and Jones (2009), and Hughen (2010). A dynamic risk premium for oil is consistent with general equilibrium

<sup>4</sup>In Baker and Routledge (2012) the dynamics are a function of risk sharing among agents while in Ready (2012) they result from supply fluctuations. Our decomposition better resembles the supply interpretation in Ready (2012) where long-run, short-run, and volatility fluctuations in supply can be priced differently in equilibrium. Ready (2013) provides evidence complementary to ours for the importance of oil supply shocks in the cross-section of returns.

<sup>5</sup>If the horizon  $T - t$  is short then non-zero correlations of futures prices with interest rates will have a negligible impact. This is the case with the horizon of implied variance data we consider.

<sup>6</sup>The return on aggregate wealth, for which the market is a proxy, appears in most equilibrium asset pricing models. Similarly, Chen, Roll, and Ross (1986), Al-Mudhaf and Goodwin (1993), Sadorsky (2001), and Sholtens and Wang (2008), among others, also include a market factor, along with proxies for oil risk factors, in their asset pricing models.

<sup>7</sup>The actual time to maturity for an oil futures contract is not precisely determined, because delivery may occur any time during the delivery month at the short side’s discretion. Following common practice, it will be assumed that delivery occurs at the end of the delivery month; this is because oil is typically in backwardation, implying that the benefit of ownership exceeds the risk-free rate and making it advantageous to hold onto the asset for as long as possible. Thus we assume the maturity date is the final business day of the delivery month.

<sup>8</sup>One can, in principle, construct several implied variance indices corresponding to different average maturities and use all of them in the estimation. In experimenting with this, we found that the term-structure of oil implied variance requires more than the single factor  $X_{v,t}$  employed in our model. Instead of expanding the model and incorporating more degrees of freedom we opted to keep the model structure the same and fit to a single implied variance series. The nearby series we employ has roughly twice the open interest and trading volume as the second-nearby.

<sup>9</sup>To best compare with Trolle and Schwartz (2009) we would have to calculate the average absolute error for implied variance. Doing so yields a goodness of fit that is comparable to theirs.

<sup>10</sup>Using two ticks, or \$0.02, for the bid-ask spread on WTI options, a conservative estimate for the bid-ask spread on the implied variance portfolio is  $2\frac{1}{\tau} \int_{F_{t,\tau}}^{\infty} \frac{\$0.02}{K^2} dK = 2\frac{\$0.02}{\tau F_{t,\tau}}$ , which averages over the sample period to roughly 1% (although it would be half as much for recent data).

<sup>11</sup>The in-sample Sharpe Ratio is obtained by taking the ratio of the daily sample mean and standard deviation of  $R_i^e$  and multiplying by the square root of 252. To help place this into perspective, the FFC portfolios (market excess return, SMB, HML, and UMD) over the same period yielded annualized in-sample Sharpe Ratios of 41%, -2%, 46%, and 54%, respectively.

<sup>12</sup>The negative  $b_v$ ’s are largely robust to changing the priors in the MCMC estimation and to the inclusion of longer maturity implied variance series (and/or more equity series). Even when overfitting to the implied variance, boosting the  $R^2$  from 98% to essentially 1, the  $b_v$ ’s remain negative though marginally significant.

<sup>13</sup>While there is much concentration in the oil industry, mostly among nationally owned companies in oil producing nations, our focus is on publicly traded international oil companies (IOCs) which arguably face a more competitive landscape. Since the OPEC price wars of 1986, in addition to competing with each other, IOCs also face curtailed profits from production and investment because of governments in consumer countries who appropriate surpluses through taxes (see Baddour, 1997).

<sup>14</sup>The VIX data are from Chicago Board Options Exchange: <http://www.cboe.com/micro/vix/historical.aspx>. We use VOX for pre-1990 period, for which CBOE does not use the new method to calculate VIX. The other macroeconomic data are from the St. Louis Fed: <http://research.stlouisfed.org/fred2/>.

<sup>15</sup>Although some of the macro variables are available at higher frequencies, we use quarterly series to better compare the forecasting power of the oil risk factors across the macro variables.

<sup>16</sup>We include the lagged macroeconomic variable in the predictive regressions as a regressor to control for the possibility that  $x_{t-1}$  predicts  $y_t$  only because  $y_{t-1}$  predicts  $y_t$  and  $x_{t-1}$  happens to be correlated with  $y_{t-1}$ . For highly persistent interest rate variables (short rate, term premium, and default premium), we use first differences ( $\Delta y_t$ ) in the left-hand-side, and levels ( $y_{t-1}$ ) in the right-hand-side.

<sup>17</sup>The relationship between  $R_{v,t}^e$  and HML is unlikely to come from having used BM-sorted oil equity portfolios in our estimation. A portfolio that is long high BM oil stocks and short low BM oil stocks has a correlation coefficient of 0.066 with the standard HML factor.

<sup>18</sup>Among SMB, HML, and UMD, the average adjusted  $R^2$  (not reported in the table) is 7.3% with HML having the largest  $R^2$  at 14.0% and UMD the smallest at 4.2%.

<sup>19</sup>The two sets of portfolios are constructed differently. The size-B/M portfolios are rebalanced once per year, while the Size-Momentum portfolios are constructed daily. This leads to small differences in the regression results when aggregating across the B/M and Momentum quintiles. In addition, aggregating over all 25 Size-B/M (or Size-Momentum) portfolio does not result in the market portfolio that we employ, and when regressed against the market and the oil factors yields a regression coefficient of 0.89 on the market along with highly significant (albeit small) oil factor coefficients.

<sup>20</sup>Because similar issues have been raised in the literature on currency risk (e.g., Dumas and Solnik, 1995; and Campbell, Serfaty-de Medeiros, and Viceira, 2010), this might be a natural candidate for an exploration using our method.

<sup>21</sup>Many other studies also restrict the risk-neutral ‘mean reversion’ matrix to have a zero eigenvalue; see Gibson and Schwartz (1990), Schwartz and Smith (2000), the USV version of the model in Hughen (2010).

**Table I. Goodness of Fit**

The mean daily root mean square errors (RMSEs) for log-futures prices are reported in basis points, the RMSEs for implied variances are reported in percentages, and  $R^2$ 's for the four portfolio returns are reported in percentages. The column labeled "Derivatives Only" lists the results using the procedure in which the model parameters and latent variables are first estimated using futures and implied variance data and then the factor loadings are estimated from equity returns. The column labeled "Full Information" lists the results estimated using futures, implied variances, and returns simultaneously. For comparison, the resulting  $R^2$ 's for the Fama-French-Carhart model are also reported.

	Derivatives Only	Full Information	FFC
Futures RMSE (basis points)	27.4	28.3	–
ImpVar RMSE (percentage)	0.8	1.7	–
Futures $R^2$ (percentage)	100.0	100.0	–
ImpVar $R^2$ (percentage)	99.80	98.1	–
$R^2$ (percentage)			
Small Oil Stocks	43.8	71.4	50.3
Large Oil Stocks	55.0	72.6	50.0
Low B/M Oil Stocks	52.5	80.3	47.7
High B/M Oil Stocks	53.8	77.1	53.7
Average $R^2$	51.3	75.3	50.4

**Table II. Subperiod and Out-of-sample Results**

The top panel reports the goodness of fit in subperiods when the model parameters and latent variables are estimated using futures, option-implied variance, and equities from the full sample. The fit to the Fama-French-Carhart model is also reported for comparison. The bottom panel reports in-sample (IS) and out-of-sample (OOS) goodness of fit in subperiods when the model parameters are estimated using data on futures, options, and equities from only one of the subperiods (IS). The mean daily root mean square errors (RMSEs) for log-futures prices are reported in basis points, RMSEs for implied variances are reported in percentages, and  $R^2$ 's for returns are reported in percentages.

Model Subsample	Full Information			Fama-French-Carhart		
	83-94	94-04	04-13	83-94	94-04	04-13
Futures RMSE (bps)	27.2	32.2	26.1	–	–	–
ImpVar RMSE (%)	1.8	1.5	2.0	–	–	–
$R^2$ (%)						
Small Oil Stocks	41.9	49.2	85.2	22.1	31.1	65.2
Large Oil Stocks	62.9	53.0	85.1	54.5	24.4	68.3
Low B/M Oil Stocks	60.4	62.3	90.8	38.2	24.8	64.2
High B/M Oil Stocks	54.8	56.6	89.4	40.6	37.2	66.8
Average	55.0	55.3	87.6	38.8	29.4	66.1

	1983 to 1994			1994 to 2004			2004 to 2013		
	83-94 (IS)	94-04 (OOS)	04-13 (OOS)	83-94 (OOS)	94-04 (IS)	04-13 (OOS)	83-94 (OOS)	94-04 (OOS)	04-13 (IS)
Futures RMSE (bps)	27.8	32.0	28.3	30.2	32.6	28.5	29.1	33.1	28.0
ImpVar RMSE (%)	1.5	1.4	1.7	1.7	1.5	1.9	1.6	1.5	1.8
$R^2$ (%)									
Small Oil Stocks	33.1	48.9	67.3	27.5	50.8	70.7	30.8	47.2	84.6
Large Oil Stocks	63.5	42.7	83.3	48.9	66.1	86.1	63.9	62.5	87.8
Low B/M Oil Stocks	53.8	56.7	76.3	37.6	71.6	90.5	53.8	66.9	91.4
High B/M Oil Stocks	49.4	53.2	76.2	41.1	60.4	86.6	49.2	57.5	89.6
Average	50.0	50.4	75.8	38.8	62.2	83.5	49.4	58.5	88.4

**Table III. Risk and Return for the Oil Factors and Equity Portfolios**

Both panels report statistics based on the full sample estimates using futures, implied variance and equity returns. The top panel reports time-series statistics for the excess returns on oil factor-mimicking portfolios:  $R_{i,t}^e$ , defined to be the  $i$ th component of  $\Delta X_{t+\Delta t} = X_{t+\Delta t} - X_t - A^* \Delta t - B^* \Delta t X_t$ . The first and third factors correspond to the transient and persistent components in oil prices. The second factor corresponds to the long-run trend in oil prices, and the last factor corresponds to the stochastic volatility of oil prices. The bottom panel reports model estimates of the coefficients from the factor decomposition of returns given by  $R_t^e = b_M R_{M,t}^e + \sum_{i=s,l,p,v} b_i R_{i,t}^e + \epsilon_t$ . The contribution of oil risk factors to the portfolio's risk premium is reported in the last column ("Net Oil RP"). Means and confidence intervals for the net oil risk premia are calculated using the posterior distribution of oil risk premia estimated from the model. A "\*" ("\*\*") corresponds to significance at the 5% (1%) level.

Statistic	$R_{s,t}^e$	$R_{l,t}^e$	$R_{p,t}^e$	$R_{v,t}^e$
Sample Mean (%)	0.001	0.004	0.058	-0.016
Sample Standard Deviation (%)	1.10	2.05	2.05	1.89
Annualized sample Sharpe Ratio (%)	1.26	2.95	44.71	-13.60
Sample 5th percentile (%)	-1.68	-3.12	-3.11	-2.15
Sample 95th percentile (%)	1.66	3.17	3.13	2.32
<u>Model Implied Unconditional:</u>				
Risk premium (annualized, %)	0.58	0.04	18.33*	-8.40*
Sharpe Ratio (annualized, %)	2.57	0.80	43.34*	-39.85*

Asset	$b_M$	$b_s$	$b_l$	$b_p$	$b_v$	Net Oil RP (%)
Small Oil Stocks	0.72**	0.67**	0.10**	0.15**	-0.33**	6.02*
Large Oil Stocks	0.76**	0.51**	0.08**	0.18**	-0.25**	5.85*
Low B/M Oil Stocks	0.83**	0.73**	0.09**	0.19**	-0.37**	7.20*
High B/M Oil Stocks	0.80**	0.58**	0.08**	0.16**	-0.32**	6.10*

**Table IV. Relationship to Macroeconomic Variables**

Panel A reports the correlations, in percentages, between model-extracted oil factor excess returns and macroeconomic variables. Panel B reports regression coefficients of regressions of macroeconomic variables on contemporaneous oil factor excess returns, and the adjusted  $R^2$ 's of the regressions. Panel C reports regression coefficients of predictive regressions of macroeconomic variables on their own lags and lagged oil factor excess returns, and the adjusted  $R^2$ 's of the predictive regressions. All regression coefficients are multiplied by 100, and all adjusted  $R^2$ 's are reported in percentages. A “\*” (“\*\*”) corresponds to significance at the 5% (1%) level. Significance is based on Newey-West standard errors with 10 lags.

*Panel A: Contemporaneous Correlations; In Percentages*

Factor	Short Rate	Term Premium	Default Premium	$\Delta(\text{VIX}^2)$	$\Delta$ Inflation	Ind. Prod. Growth	$\Delta$ Unemployment	GDP Growth
$R_s^e$	-0.4	-0.3	-15.4	-12.0	24.3**	1.0	-16.0	14.3
$R_l^e$	-15.3	-2.3	5.6	-18.0	-3.6	1.9	-6.4	-7.7
$R_p^e$	6.7	3.4	-9.2	6.3	16.0	-7.9	4.7	6.9
$R_v^e$	-13.7	-1.8	54.1**	16.7	-31.7**	-4.7	37.4**	-39.8**

*Panel B: Contemporaneous Regressions; Coefficients  $\times 100$*

Variable	Constant	$R_s^e$	$R_l^e$	$R_p^e$	$R_v^e$	Adj $R^2$ (%)
Short Rate	4.1**	-0.6	-5.2**	-5.3**	-2.2*	4.9
Term Premium	1.7**	0.1	0.2	0.3	0.0	-3.7
Default Premium	1.0**	0.1	0.2	0.2	1.0**	26.9
$\Delta(\text{VIX}^2)$	0.4	-0.9	-9.3**	-9.5*	2.2	7.1
$\Delta$ Inflation	-0.6*	8.1**	7.8**	13.8**	-2.5	19.3
Ind. Prod. Growth	0.0	-2.9	-7.4*	-11.3*	-3.3*	-0.1
$\Delta$ Unemp. Rate	0.0	0.0	0.1	0.2	0.6**	12.0
GDP Growth	2.9**	0.4	-2.6*	-2.5	-4.6**	14.4

*Panel C: Predictive Regressions at 1 Quarter Lag; Coefficients  $\times 100$*

Variable	Constant	$R_s^e$	$R_l^e$	$R_p^e$	$R_v^e$	Own Lag	Adj $R^2$ (%)
Short Rate	0.0	-0.2	0.1	0.1	-0.2	-1.9	-3.3
Term Premium	0.2*	-0.4	0.2	0.3	0.1	-10.7**	3.1
Default Premium	0.3*	-0.1	-0.1	-0.1	0.2	-26.3**	8.8
$\Delta(\text{VIX}^2)$	-0.1	-0.2	2.8	1.4	0.3	-35.4**	11.6
$\Delta$ Inflation	-0.1	-1.6	0.7	0.3	1.8	-41.8*	19.0
Ind. Prod. Growth	0.1	-2.4	-8.9	-14.7	-11.3**	17.2	18.3
$\Delta$ Unemp. Rate	0.0	0.1	0.2	0.2	0.4**	46.8**	33.5
GDP Growth	1.6**	-2.0*	-1.7	-3.3	-2.9**	43.3**	28.8

**Table V. Relationship between Fama-French-Carhart and Oil Factors**

Panel A reports the coefficients from regressing each oil factor excess return on the Fama-French-Carhart factors (SMB, HML, and UMD) and the market's excess returns (EMKT). Panel B reports the coefficients from regressing each of the FFC factors on the oil factors and market excess returns. The results of a regression of the market on the oil factors excess returns are also reported. Partial Oil  $R^2$  are calculated as the difference between the regression adjusted  $R^2$  and the adjusted  $R^2$  from a regression employing only market excess returns as a regressor. The oil betas,  $\beta_j$  are the regression coefficients normalized so that the Fama-French oil industry ("FF oil ind") is defined to have  $(\beta_s, \beta_l, \beta_p, \beta_v) = (1, 1, 1, -1)$ . Means and confidence intervals for the annualized net oil risk premia (reported in the last column) are calculated using the posterior distribution of oil risk premia estimated from the model. A "\*" (\*) corresponds to significance at the 5% (1%) level.

*Panel A: Oil Factor Decomposition*

Dependent Variable	EMKT	SMB	HML	UMD	Alpha (%)	Adj $R^2$
$R_s^e$	0.12**	0.39**	0.44**	0.02	-0.010	7.8
$R_l^e$	0.05*	0.27**	0.32**	0.14**	-0.006	1.1
$R_p^e$	0.13**	-0.18**	0.03	-0.05	0.055*	0.9
$R_v^e$	-0.47**	-0.58**	-0.91**	0.02	0.011	12.6

*Panel B: FFC Factor Decomposition*

Dependent Variable	Partial Oil						Net Oil				
	$b_M$	$b_s$	$b_l$	$b_p$	$b_v$	$R^2$ (%)	$\beta_s$	$\beta_l$	$\beta_p$	$\beta_v$	RP (%)
EMKT	NA	0.03**	0.01	0.05**	-0.11**	NA	NA	NA	NA	NA	NA
FF oil ind	0.76**	0.45**	0.07**	0.17**	-0.21**	23.6	1	1	1	-1	5.34**
SMB	-0.09**	0.08**	0.02**	0.00	-0.04**	4.9	0.19	0.21	0.00	-0.18	0.40
HML	-0.14**	0.09**	0.03**	0.03**	-0.06**	8.8	0.20	0.36	0.15	-0.28	1.05*
UMD	-0.12**	-0.03**	0.01*	0.00	0.03**	0.9	-0.07	0.20	0.00	0.16	-0.32

**Table VI. Portfolio Oil Betas and Annualized Percentage Risk Premia for Each Fama-French Size-Characteristic Sorted Portfolio**

The characteristics used are book-value (“Book”) and momentum (“Mom”) and the data is taken from Kenneth French’s website. The full-sample returns on each portfolio are regressed against the excess oil returns and the market to calculate oil betas,  $\beta_j$ . The  $\beta_j$ ’s are normalized so that the Fama-French oil industry is defined to have  $(\beta_s, \beta_t, \beta_p, \beta_v) = (1, 1, 1, -1)$ . Means and confidence intervals for the net oil risk premia (reported in the bottom panel) are calculated using the posterior distribution of oil risk premia estimated from the model. A “\*\*” (“\*\*\*”) corresponds to significance at the 5% (1%) level.

	Book	Size					Mom	Size				
		1	2	3	4	5		1	2	3	4	5
$\beta_s$	1	0.25**	0.15**	0.08**	-0.01	-0.19**	1	0.35**	0.29**	0.27**	0.23**	0.06*
	2	0.25**	0.22**	0.16**	0.16**	-0.01	2	0.31**	0.25**	0.20**	0.16**	0.00
	3	0.25**	0.27**	0.23**	0.23**	0.12**	3	0.28**	0.24**	0.21**	0.15**	0.00
	4	0.27**	0.24**	0.24**	0.24**	0.07**	4	0.28**	0.26**	0.20**	0.15**	-0.02
	5	0.30**	0.31**	0.21**	0.18**	0.09**	5	0.29**	0.25**	0.22**	0.15**	-0.02
$\beta_t$	1	0.20**	0.10	0.00	-0.10	-0.22**	1	0.30**	0.16	0.02	0.16	-0.06
	2	0.24**	0.24**	0.19**	0.19**	-0.03	2	0.35**	0.27**	0.21**	0.06	-0.07
	3	0.33**	0.36**	0.39**	0.30**	0.17**	3	0.35**	0.37**	0.30**	0.14**	0.00
	4	0.36**	0.37**	0.43**	0.31**	0.10	4	0.38**	0.36**	0.31**	0.23**	0.08
	5	0.40**	0.50**	0.33**	0.27**	0.29**	5	0.47**	0.42**	0.35**	0.31**	0.20**
$\beta_p$	1	0.01	-0.05	-0.02	-0.03	-0.12**	1	0.05	0.04	0.07	0.13**	0.10*
	2	0.01	0.00	0.01	0.05**	-0.01	2	0.06*	0.03	0.06*	0.02	-0.03
	3	0.04	0.04	0.09**	0.12**	0.07**	3	0.05*	0.06*	0.07**	0.03*	0.00
	4	0.06*	0.06*	0.11**	0.13**	0.07**	4	0.07**	0.03	0.05**	0.05**	0.00
	5	0.09**	0.07*	0.07**	0.06*	0.09**	5	0.09**	0.05	0.08**	0.10**	0.05*
$\beta_v$	1	-0.32**	-0.13**	-0.05**	0.08**	0.19**	1	-0.49**	-0.39**	-0.30**	-0.24**	-0.24**
	2	-0.33**	-0.25**	-0.15**	-0.21**	-0.04**	2	-0.48**	-0.38**	-0.33**	-0.30**	-0.09**
	3	-0.35**	-0.29**	-0.25**	-0.32**	-0.15**	3	-0.43**	-0.35**	-0.30**	-0.26**	-0.11**
	4	-0.34**	-0.33**	-0.34**	-0.33**	-0.21**	4	-0.39**	-0.32**	-0.25**	-0.19**	-0.03**
	5	-0.39**	-0.36**	-0.25**	-0.30**	-0.14**	5	-0.34**	-0.28**	-0.18**	-0.09**	0.11**
Net Oil	1	0.70	0.10	0.05	<b>-0.24*</b>	<b>-0.80*</b>	1	1.16	0.92	0.85	<b>0.93*</b>	<b>0.77**</b>
	2	0.71	0.52	0.37	0.59	0.03	2	1.17	0.89	0.85	0.67	0.08
RP (%)	3	0.85	0.76	0.81	<b>1.04*</b>	<b>0.53*</b>	3	1.04	0.89	0.84	0.63	0.21
	4	0.89	0.86	1.05	<b>1.09*</b>	<b>0.60*</b>	4	1.01	0.77	0.69	0.57	0.06
	5	1.10	0.99	0.73	0.81	0.57	5	1.00	0.75	0.66	0.53	-0.06

**Table VII. Industries Most Frequently Represented in the Top and Bottom Octiles, Sorted Based on Exposure to the Oil Factors**

Each day, the excess returns of each of the 48 non-oil industries based on the classification in Kenneth French’s website are regressed on the four oil factors’ excess returns and the market, using the previous 60 trading days’ returns. The 48 industries can then be sorted into eight octiles according to the regression coefficient on  $R_{i,t}^e$ : The six industries with the highest (lowest) recent past sensitivity to  $R_{i,t}^e$  are equally weighted at date- $t$  for the purpose of calculating the date- $t$  return of the eighth (first) octile sorted by sensitivity to  $X_i$ .

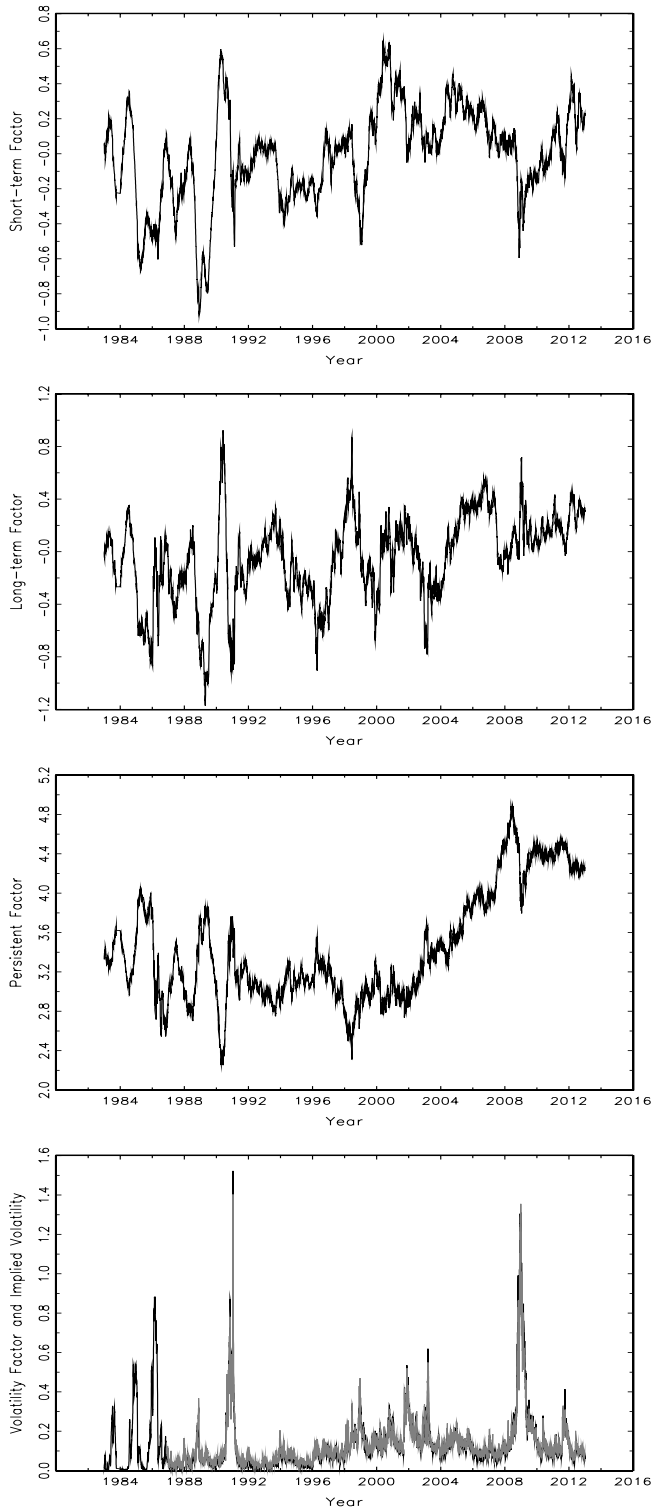
		Sorting by Sensitivity to							
		$X_s$		$X_l$		$X_p$		$X_v$	
	Industry	Freq (%)	Industry	Freq (%)	Industry	Freq (%)	Industry	Freq (%)	
Octile 1	Smoke	5.02	Gold	4.56	Smoke	3.91	Coal	8.76	
	Beer	4.22	Soda	4.49	Agric	3.48	Gold	8.16	
	Hardw	4.19	Smoke	4.38	Guns	3.44	Mines	6.17	
	Soda	4.17	Hardw	3.82	Autos	3.3	Steel	5.22	
	Guns	3.73	Chips	3.48	Soda	3.26	Cnstr	4.98	
Octile 8	Gold	10.9	Coal	6.39	Gold	10.15	Hardw	5.61	
	Coal	9.6	Gold	6.25	Coal	8.22	Softw	4.61	
	Mines	8.19	FabPr	4.46	Mines	6.34	Soda	4.54	
	Cnstr	5.35	Mines	4.24	Cnstr	5.06	Chips	4.08	
	Steel	5.19	Ships	3.85	Agric	3.81	Smoke	3.98	

Short Name	Industry
Agric	Agriculture
Autos	Automobiles and Trucks
Beer	Beer & Liquor
Soda	Candy & Soda
Coal	Coal
Softw	Computer Software
Hardw	Computers
Cnstr	Construction
Guns	Defense
Chips	Electronic Equipment
FabPr	Fabricated Products
Mines	Non-Metallic and Industrial Metal Mining
Gold	Precious Metals
Ships	Shipbuilding, Railroad Equipment
Steel	Steel Works Etc
Smoke	Tobacco Products

**Table VIII. Portfolio Oil Betas and Annualized Percentage Risk Premia for Industries Sorted on Sensitivity to Oil Factors**

Each day, the excess returns of each of the 48 non-oil industries based on the classification in Kenneth French's website are regressed on the four oil factors' excess returns and the market, using the previous 60 trading days' returns. The 48 industries can then be sorted into eight octiles according to the regression coefficient on  $R_{i,t}^e$ : The six industries with the highest (lowest) recent past sensitivity to  $R_{i,t}^e$  are equally weighted at date- $t$  for the purpose of calculating the date- $t$  return of the eighth (first) octile sorted by sensitivity to  $X_i$ . This is done daily for each of the  $X_i$ 's to produce 32 octile portfolios. The full-sample returns on each octile portfolio is then regressed against the excess oil returns and the market to calculate its oil betas,  $\beta_j$ . To facilitate comparisons, we normalize the  $\beta_j$ 's so that the Fama-French oil industry is defined to have  $(\beta_s, \beta_l, \beta_p, \beta_v) = (1, 1, 1, -1)$ . Means and confidence intervals for the net oil risk premia (reported in the bottom panel) are calculated using the posterior distribution of oil risk premia estimated from the model. A “\*” (“\*\*”) corresponds to significance at the 5% (1%) level.

Sorted by		Sensitivity Octile								
		1	2	3	4	5	6	7	8 minus 1	
$\beta_s$	$X_s$	-0.07**	-0.02*	-0.01	0.00	0.06**	0.08**	0.17**	0.57**	0.65**
	$X_l$	0.03	0.02	0.04**	0.02	0.04**	0.08**	0.17**	0.39**	0.37**
	$X_p$	0.01	0.00	0.02*	0.04**	0.05**	0.09**	0.12**	0.45**	0.45**
	$X_v$	0.51**	0.18**	0.10**	0.07**	0.04**	-0.02*	-0.04*	-0.06**	-0.57**
$\beta_l$	$X_s$	-0.04	0.06	0.03	0.07	0.07	0.17**	0.27**	0.75**	0.79**
	$X_l$	0.00	0.08	0.03	0.11*	0.11*	0.11*	0.29**	0.64**	0.64**
	$X_p$	0.16**	0.00	0.11*	0.10*	0.02	0.17**	0.19**	0.62**	0.46**
	$X_v$	0.66**	0.27**	0.26**	0.15**	0.12**	-0.01	0.06	-0.13*	-0.80**
$\beta_p$	$X_s$	-0.10**	-0.08**	-0.10**	-0.06**	-0.06**	-0.01	0.08**	0.47**	0.58**
	$X_l$	-0.01	-0.05**	-0.05**	-0.04*	-0.07**	-0.04*	0.09**	0.33**	0.35**
	$X_p$	-0.03	-0.08**	-0.06**	-0.09**	-0.07**	0.00	0.06**	0.42**	0.46**
	$X_v$	0.41**	0.07**	0.00	-0.04*	-0.04*	-0.12**	-0.07**	-0.08**	-0.49**
$\beta_v$	$X_s$	0.05**	0.02	-0.07**	-0.09**	-0.12**	-0.18**	-0.29**	-0.63**	-0.68**
	$X_l$	0.00	-0.04**	-0.09**	-0.12**	-0.11**	-0.19**	-0.26**	-0.48**	-0.48**
	$X_p$	-0.02	-0.03**	-0.10**	-0.11**	-0.15**	-0.18**	-0.19**	-0.51**	-0.50**
	$X_v$	-0.65**	-0.33**	-0.18**	-0.13**	-0.07**	-0.05**	0.02	0.08**	0.73**
Net Oil	$X_s$	-0.47**	-0.29*	-0.21	-0.03	0.03	0.31	0.82*	2.84*	3.29**
RP (%)	$X_l$	-0.05	-0.10	0.00	0.09	-0.02	0.22	0.80*	2.06*	2.11*
	$X_p$	-0.08	-0.21	-0.02	-0.09	0.05	0.35	0.57*	2.44**	2.52**
	$X_v$	2.66*	0.86	0.37	0.12	0.00	-0.32*	-0.27**	-0.42*	-3.09**



**Figure 1. Estimated Oil Factors and Option Implied Variance**

This figure depicts time series of estimated latent oil risk factors at the mean of their posterior distributions: The transient component,  $X_{s,t}$  (top panel); the long-term component,  $X_{l,t}$  (second panel); the persistent component,  $X_{p,t}$  (third panel); and the unspanned volatility  $X_{v,t}$  (bottom panel, in black and plotted together with option-implied variance in grey).