CFS WORKING PAPER

No. 466

A General Approach to Recovering Market Expectations from Futures Prices with an Application to Crude Oil

Christiane Baumeister, Lutz Kilian

Center for Financial Studies
Goethe University ■ House of Finance
Grüneburgplatz 1 ■ 60323 Frankfurt am Main

Tel: +49 69 798-30050
Fax: +49 69 798-30077
www.ifk-cfs.de ■ info@ifk-cfs.de

CFS Working Paper Series

The *Center for Financial Studies*, located in Goethe University's House of Finance in Frankfurt, is an independent non-profit research center, funded by the non-profit- making organisation Gesellschaft für Kapitalmarktforschung e.V. (GfK). The CFS is financed by donations and by contributions of the GfK members, as well as by national and international research grants. The GfK members comprise major players in Ger- many's financial industry. Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. CFS is also a contributor to policy debates and policy analyses, building upon relevant findings in its research areas.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center´s Research Fellow Program or members of one of the Center´s Research Projects.

If you would like to know more about the Center for Financial Studies, please let us know of your interest.

Prof. Michalis Haliassos, Ph.D. Prof. Dr. Jan Pieter Krahnen Prof. Dr. Uwe Walz

Center for Financial Studies
Goethe University ■ House of Finance
Grüneburgplatz 1 ■ 60323 Frankfurt am Main

Tel: +49 69 798-30050
Fax: +49 69 798-30077
www.ifk-cfs.de ■ info@ifk-cfs.de

A General Approach to Recovering Market Expectations from Futures Prices With an Application to Crude Oil

September 13, 2014

Christiane Baumeister
Bank of Canada

Lutz Kilian
University of Michigan

Futures markets are a potentially valuable source of information about market expectations. Exploiting this information has proved difficult in practice, because the presence of a time-varying risk premium often renders the futures price a poor measure of the market expectation of the price of the underlying asset. Even though the expectation in principle may be recovered by adjusting the futures price by the estimated risk premium, a common problem in applied work is that there are as many measures of market expectations as there are estimates of the risk premium. We propose a general solution to this problem that allows us to uniquely pin down the best possible estimate of the market expectation for any set of risk premium estimates. We illustrate this approach by solving the long-standing problem of how to recover the market expectation of the price of crude oil. We provide a new measure of oil price expectations that is considerably more accurate than the alternatives and more economically plausible. We discuss implications of our analysis for the estimation of economic models of energy-intensive durables, for the debate on speculation in oil markets, and for oil price forecasting.

JEL Code: C53, D84, G14, Q43
Key Words: Futures, risk premium, market expectation, model uncertainty, forecast, oil price.

Acknowledgements: The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Bank of Canada. We thank David Finer for excellent research assistance. We especially thank Cynthia Wu for providing access to the code used in Hamilton and Wu (2014a). We have also benefited from helpful discussions with Ryan Kellogg, Stefan Nagel, and Cynthia Wu.

Corresponding author: Lutz Kilian, Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA. Email: lkilian@umich.edu.

Christiane Baumeister, International Economic Analysis Department, Bank of Canada, 234 Laurier Avenue West, Ottawa, ON K1A 0G9, Canada. Email: CBaumeister@bankofcanada.ca.

1. Introduction

Futures markets are a potentially valuable source of information about market expectations. Exploiting this information has proved difficult in practice, because the presence of a time-varying risk premium often renders the futures price a poor measure of the market expectation of the price of the underlying asset. Even though the market expectation in principle may be recovered by adjusting the futures price by the estimated risk premium, a common problem in applied work is that there are as many measures of market expectations as there are estimates of the risk premium, and these risk premium estimates may differ substantially. Thus, attempts to pin down the market expectation have often proved elusive. We propose a general solution to this problem that allows us to uniquely pin down the best possible estimate of the market expectation for any set of risk premium estimates. The central idea is that – in the presence of a risk premium – the risk-adjusted futures price is the conditional expectation of the price and hence the minimum mean-squared prediction error (MSPE) predictor by construction (see Granger 1969). This fact allows one to rank alternative model specifications based on their MSPE and to identify the most accurate measure of expectations.

We illustrate this approach by solving the long-standing problem of how to recover the market expectation of the price of crude oil. For this purpose, we provide a comprehensive analysis of the time-varying risk premium in the oil futures market. We show that alternative estimates of the risk premium for the same month may differ by as much as \$56. We demonstrate that the expectations measure selected by our procedure is a considerably more accurate measure of oil price expectations than the unadjusted futures price or any other risk-adjusted futures price. Using this approach also generates an economically more plausible path of expectations than alternative risk adjustments.

Our analysis helps explain the apparent failure of the oil futures price as a predictor of the oil price during the surge in the spot price of oil between 2003 and mid-2008. It provides an improved measure of oil price expectations that can be used in estimating economic models of automobile purchases, investment decisions under uncertainty, environmental policies and regulatory reforms. We provide monthly time series estimates of the market expectation of the price of oil since 1992. We show, for example, that since 2010 the one-year-ahead market expectation of the price of oil has stabilized near $90.

It may seem that the problem of identifying the market expectation could alternatively have been solved by searching for the model with the most predictive power for the excess return on oil futures contracts. Indeed, this is one metric by which return regressions in the literature have often been evaluated. There is no reason, however, for the model that minimizes the MSPE for the excess rate of return also to minimize the MSPE for the spot price of oil expressed in dollars because the loss functions differ. In fact, it can be shown that minimizing the MSPE of the rate of return produces inaccurate measures of oil price expectations. In addition, evaluating the risk premium models under a different loss function than the loss function used in their estimation also helps deal with the problem of data mining in fitting return regressions.

The remainder of the paper is organized as follows. Section 2 provides some background on the use of oil futures prices in the literature, it reviews the emerging consensus that there is a time-varying risk premium in oil futures markets, and it motivates our approach to discriminating between alternative risk premium estimates. In section 3 we establish notation and characterize the excess return data in the oil futures market. Section 4 briefly reviews the candidate models of the time-varying risk premium proposed in the literature. In section 5 we estimate these models and document that the risk premium estimates differ greatly across alternative model specifications. In section 6 we evaluate the merits of these estimates based on the predictive content of the implied oil price expectations. Section 7 introduces a generalized return regression that encompasses all predictors used in the candidate models and is more accurate than all return regressions proposed in the literature. In Section 8 we assess the economic plausibility of the two most successful model specifications and show that one measure of oil price expectations dominates all others. Section 9 examines the out-of-sample accuracy of real-time risk-adjusted forecasts. Section 10 summarizes additional results based on quarterly data contained in the not-for-publication appendix. The concluding remarks are in section 11.

2. Risk premia in the oil futures market: What we know and why we care

The price of oil is one of the key economic variables for the assessment of macroeconomic performance and risks at central banks and international organizations. It plays an important role in designing environmental policies, and it has an immediate impact on a wide range of industries such as the automobile industry, airlines, and utility companies. It also has implications for the economic viability of the production of crude oil from Canadian oil sands and the viability of U.S. shale oil production, which directly affects the energy security of the United States. The evolution of the price of oil is highly uncertain and difficult to predict with a reasonable degree of accuracy. For many years, the standard practice among policymakers and central bankers, in the business community, in the financial press and in the academic literature, has been to interpret the price of West Texas Intermediate (WTI) crude oil futures as the market expectation of the spot price of WTI crude oil. The use of oil futures prices as out-of-sample oil price forecasts relies on this interpretation, as does the use of oil futures prices as a measure of oil price expectations of firms and consumers in microeconomic models.

The popularity of this approach has several reasons. First, futures prices are simple to use and readily available in real time. Second, there is a reluctance to depart from what is viewed as the collective wisdom of the financial market which presumably knows better than any individual oil price forecaster. Relying on what is perceived to be the market expectation also absolves the forecaster from any culpability for forecast errors because no one can reasonably be expected to beat the market. Third, there is evidence that futures prices have outperformed other oil price forecasts on average at least at some forecast horizons, although their forecast accuracy has varied substantially over time. Moreover, until recently there were few alternatives available to oil price forecasters (e.g., Alquist, Kilian and Vigfusson 2013). Fourth, while it is well understood that time-varying risk premia would invalidate the use of oil futures prices as oil price forecasts, it has proved difficult to reject the absence of a time-varying risk premium based on the traditional statistical tests of forecast efficiency proposed by Fama and French (1987, 1988) (see, e.g., Alquist and Kilian 2010).

This practice has been challenged in recent years by a large number of empirical studies documenting the existence of time-varying risk premia in the oil futures market. Examples include De Roon, Nijman, and Veld (2000), Sadorsky (2002), Pagano and Pisani (2009), Acharya, Lochstoer, and Ramadorai (2013), Etula (2013), and Singleton (2014). These studies move beyond the statistical framework proposed by Fama and French (1987, 1988). They provide direct evidence that excess returns in oil futures markets can be predicted using a range of aggregate and commodity-market specific financial and macroeconomic variables. A new consensus has been emerging in the academic literature that time-varying risk premia are an important feature of the crude oil market. For example, Singleton (2014) concludes that "the evidence for time-varying risk premiums in oil markets … seems compelling".

The possible presence of a time-varying risk premium in oil futures markets has potentially far-reaching implications for modeling energy price expectations in economic models, for the policy debate about the role of speculation in oil markets, and for oil price forecasters. First, there is a long tradition of using oil futures prices as proxies for energy price expectations in empirical models of the purchases of energy-intensive durables, in models of the effect of uncertainty on investment decisions, and in models of the impact of regulatory policies such as automotive fuel standards and gasoline taxes (e.g., Busse, Knittel and Zettelmeyer 2013; Kellogg 2014; Allcott and Wozny 2014). This practice is questionable because in the presence of a risk premium the oil futures price differs from the expectation of the price of oil. Being able to control for variation in the risk premium hence allows one to construct improved measures of the market expectations of the price of oil.

Second, obtaining improved measures of market expectations allows one to understand better the evolution of the price of crude oil. For example, it has been observed that the term structure of futures prices remained largely flat throughout much of the 2003-08 period, even though realized oil prices increased persistently. This pattern is potentially consistent with a risk premium being priced in already, but it also is consistent with the market being repeatedly surprised by these oil price increases. Knowledge of the time-varying risk premium helps us separate these explanations.

Third, recent structural models of speculation in the physical market for oil rely on the assumption of a risk premium that is zero or at least constant over time. Estimates of these models suggest that recent oil price fluctuations cannot be explained as the result of shifts in speculative demand, removing one of the key rationales for recent efforts to tighten the regulation of oil derivatives markets (e.g., Fattouh et al. 2013; Kilian and Murphy 2014; Kilian and Lee 2014). The presence of a large time-varying risk premium could potentially undermine the validity of these models with important implications for regulators.

Finally, oil futures prices are widely used as oil price forecasts by practitioners. The existence of a time-varying risk premium raises the question of whether the accuracy of out-of-sample oil price forecasts may be improved by adjusting the oil futures price by real-time estimates of the risk premium. Not only are more accurate oil price forecasts likely to improve the accuracy of forecasts of macroeconomic outcomes, but many industries depend directly on forecasts of the price of oil.[1]

Our paper provides a systematic investigation of the evidence for time-varying risk premia in the crude oil market. Although the evidence compiled in the existing literature may seem overwhelming at first sight, closer inspection reveals that it is difficult to draw general conclusions from the empirical studies in question because they differ along many dimensions. First, they differ greatly in the sample period covered and in the horizon for which the risk premium is computed. Moreover, often they evaluate the predictor of interest over very short time periods only. This is a particular concern given the well-documented instability of predictive relationships in oil markets. For example, Hamilton and Wu (2014b) document that the empirical results in Singleton (2014) are not robust to extending the sample by only a few years.

Second, sometimes in this debate little distinction is made between results for crude oil and for other commodities. For example, Singleton (2014) cites Fama and French (1987) as having provided evidence of a time-varying risk premium in the crude oil market, yet oil was

[1] For example, airlines rely on such forecasts in setting airfares, automobile companies decide their product menu and set product prices with oil price forecasts in mind, and utility companies use oil price forecasts in deciding whether to expand capacity or to build new plants. Likewise, homeowners rely on oil price forecasts in deciding the timing of their heating oil purchases or whether to invest in energy-saving home improvements. Forecasts of the price of oil also play an important role in generating projections of energy use, in predicting carbon emissions and climate change, and in assessing the macroeconomic impacts of oil price shocks.

never considered in their paper. In fact, at the time Fama and French (1987) was published, the WTI oil futures market was still in its infancy. Other studies commonly cited in the debate about time-varying risk premia in oil futures prices that, in fact, do not analyze the futures market for crude oil include Bessembinder and Chan (1992), Chong and Miffre (2010), and Cheng, Kirilenko and Xiong (2012).

Third, a closely related problem is that many studies do not explicitly focus on crude oil, but estimate time-varying risk premia for portfolios of several energy commodities including, for example, natural gas along with crude oil. This portfolio approach is problematic because the wellhead price of natural gas in recent years fell dramatically, while the price of crude oil surged. Likewise, studies of portfolios including refined products such as gasoline or heating oil along with crude oil are not informative about the question of time-varying risk premia in the crude oil market. A case in point is the study by Hong and Yogo (2012), which provides results for an energy portfolio consisting of heating oil, gasoline, crude oil, natural gas, and propane gas, but no results that are directly relevant for the crude oil market.

Finally, a striking feature of this literature is that there is general agreement that the risk premium is time-varying, but no agreement as to which predictors have the most predictive power for excess returns. Whereas one study might favor one set of predictors, the next study may focus on an entirely different set of predictors. This fact not only raises questions about the economic plausibility of these models, but also suggests that these results may be subject to data mining biases (e.g., Inoue and Kilian 2004). Moreover, there is little agreement on how to measure the predictive success of a given model. In many cases, researchers have focused exclusively on the question of whether the *t*-statistics of the preferred predictor of excess returns is statistically significant at conventional significance levels (possibly after conditioning on other predictors) with no consideration for the magnitude of the estimated risk premium.

Our objective in this paper is to provide a systematic evaluation of the predictive power of a range of predictors of excess returns proposed in this literature. We quantify the estimated risk premia in dollar terms and investigate their sign, their magnitude and their variability across alternative model specifications. We focus on the most influential and most widely cited studies in this literature. In some cases, when the original study did not consider crude oil specifically we extend the analysis to the oil futures market. For example, we report additional results for the model specification originally considered in Bessembinder and Chan (1992) for other commodities. We also consider predictive models that heretofore have been applied only to portfolios rather than to crude oil specifically. For each predictive model underlying the emerging consensus that there are time-varying risk premia in oil futures markets, we follow as closely as possible the data definitions and model specifications proposed in the literature, but we extend the sample until June 2014, which in many cases amounts to a substantial increase in the sample size. We also consider prediction horizons of 3, 6, 9, and 12 months for all models.[2] Finally, we make sure to evaluate all model specifications based on the same evaluation period to the extent that the data are available.

We document that there is tremendous variability in the estimates of the time-varying risk premium across model specifications. This variability raises concerns about the validity of these estimates. It also creates a further challenge when measuring oil price expectations. While the idea of measuring market expectations by risk adjusting futures prices has a long tradition in the literature, different estimates of the risk premium imply very different estimates of the market

[2] Restricting the horizon to 12 months allows a systematic evaluation of the accuracy of the predictive models over a long evaluation period. Clearly, for some applications longer maturities are of interest. For example, Kellogg (2014) focuses on the 18-month futures contract. It would be straightforward to extend our analysis to such horizons, but only at the cost of a shorter evaluation sample.

expectation of the oil price, so it is not clear how to proceed in practice. This situation is not unique to oil futures markets. It is a common feature of many futures markets. We show that this model uncertainty can be resolved by appealing to the theoretical result that – in the presence of a risk premium – the risk-adjusted futures price is the conditional expectation of the price and hence the minimum MSPE predictor by construction (see Granger 1969). This fact suggests that after adjusting the futures price for the estimated risk premium, the MSPE of the forecast should decline. If it does not, then the estimate of the risk premium is not credible. Moreover, the most plausible measure of oil price expectations, given a set of candidate models, is the specification that produces the minimum MSPE. Below we illustrate this general idea in the context of the problem of recovering expectations about the future path of the spot price of crude oil from oil futures prices.

3. Excess returns in oil futures markets

Oil futures markets facilitate the transfer of risk to market participants who are willing to bear it. The market price of this risk is known as the risk premium and equals the difference between the current oil futures price and the expected spot price of crude oil at the delivery date. The literature has focused on two main determinants of the risk premium. One is oil-specific risk. The other determinant is systematic risk, defined as risk that cannot be diversified away. The magnitude of the premium for systematic risk depends on the covariance between oil futures prices and changes in the state of the economy. There is no general agreement on the relative importance of systematic and oil-specific determinants of the risk premium and no theoretical framework that encompasses all explanations.

In the absence of a time-varying risk premium, one would expect the forecast errors from using the oil futures price as a predictor of the oil price to be uncorrelated with any variable in the information set of the forecaster. Let F_t^h denote the current dollar price of an oil futures contract maturing h periods from now, and S_{t+h} the corresponding spot price of oil at the delivery date of the futures contract. The random payoff of a long futures position is $S_{t+h} - F_t^h$. Standard no arbitrage arguments imply that

$$E_t\left[Q_{t+h}\left(S_{t+h} - F_t^h\right)\right] = 0,$$

where Q_{t+h} denotes a stochastic pricing kernel. Rearranging this expression yields

$$F_t^h = E_t\left[S_{t+h}\right] + \text{cov}\left(F_t^h, Q_{t+h}\right) / E_t\left[Q_{t+h}\right]$$

where $\text{cov}\left(F_t^h, Q_{t+h}\right) / E_t\left[Q_{t+h}\right]$ refers to the risk premium. It is readily apparent that $F_t^h = E_t\left[S_{t+h}\right]$ if and only if $\text{cov}\left(F_t^h, Q_{t+h}\right) / E_t\left[Q_{t+h}\right] = 0.$ In that case, $E_t\left[F_t^h - S_{t+h}\right] = 0,$ where $F_t^h - S_{t+h}$ denotes the error from predicting the price of oil based on the oil futures price. This prediction error also is the return on futures because $F_t^h - F_{t+h}^{t+h} = F_t^h - S_{t+h}$. Because positions in futures markets do not require investment outlays, these returns may equivalently be referred to as excess returns (see Etula 2013, p. 13). These terminologies are used interchangeably in the literature. The reference to excess returns emphasizes the parallels between the analysis of time-varying risk premia in oil futures markets and in other futures markets (e.g., Piazzesi and Swanson 2008).

Evidence of a predictable component in the prediction error such that $E_t\left[F_t^h - S_{t+h}\right] \neq 0$ would be consistent with the presence of a time-varying risk premium. Of course, such evidence also might be explained by a model of rational learning in the oil futures market in the absence of a risk premium (e.g., Timmermann 1993). Yet another explanation of this pattern could be that traders priced in a possible collapse of oil prices that was expected to occur with low probability, but failed to materialize within this period. We follow the literature in abstracting from these alternative explanations. We examine instead the extent to which a time-varying risk premium can explain the forecast errors implied by forecasts based on oil futures prices. We do not presume, however, that a time-varying risk premium alone can account for these forecasts errors. Rather our analysis should be viewed as complementary to other potential explanations. Indeed, to the extent that accounting for the risk premium improves our understanding of the underlying market expectation, the case can be made that our analysis helps provide the empirical basis for evaluating models of learning and of peso problems in oil markets.

To facilitate a more formal analysis of the forecast errors, Figure 1 plots the monthly excess returns. For expository purposes, we focus on the 3-month and 12-month horizon. The plot shows the returns starting from February and November 1989, respectively, to June 2014 with each point in the plot depicting the realized excess return $\left(F_t^h - S_{t+h}\right)/S_{t+h}$ at date $t+h.$ All results are expressed in percent deviations. On average over the sample the excess returns are slightly negative. At the 12-month horizon, for example, they range from -$54 to $137 with a standard deviation of $30 about a mean of -$5. While the excess return appears stationary, there is evidence of serial correlation in the excess returns even at the 3-month horizon. There also is evidence that the peaks and troughs are not random. For example, in the 12-month returns there is a sharp spike 12 months after demand for oil began to weaken in early 2008. There is also a trough exactly 12 months after the all-time low in the WTI spot price of oil in February 1999. For many spikes and troughs in Figure 1 the economic interpretation is somewhat ambiguous, however. Next we assess whether fluctuations in the excess return are predictable based on publicly available information.

4. Empirical models of time-varying risk premia in oil futures markets

Evidence that oil futures prices are unbiased predictors of the spot price of oil implies that the average risk premium is zero. We are not the first to study the average risk premium in oil futures markets (see, e.g., Chernenko et al. 2004, Pagano and Pisani 2009, Alquist and Kilian 2010). Reexamining this question is nevertheless useful, as earlier studies in some cases obtained conflicting results. Assessing the statistical significance of the average risk premium is complicated by the fact that $F_t^h - S_{t+h}$ is clearly nonstationary in that its variance declines over time. Although as economists we are interested in the average risk premium measured in dollars, evaluating the statistical significance of the mean of $F_t^h - S_{t+h}$ requires data that are stationary. A common approach is to express the return in percent changes such that

$$\left(F_t^h - S_{t+h}\right) / S_{t+h} = \alpha + e_{t+h},$$

where e_{t+h} is the serially correlated and possibly heteroskedastic regression error. Then a test of the null hypothesis of an average risk premium of zero can be conducted as a two-sided t-test of $H_0 : \alpha = 0.$ Table 1 shows there is no evidence of a statistically significant average risk premium at any horizon. Even more important than the question of statistical significance is the question of whether the average risk premia are large enough to substantively affect the analysis of oil futures markets. The last column of Table 1 shows that average risk premia expressed in dollars are all economically insignificant. The risk premium ranges from -73 cents to -3.51 dollars, depending on the horizon. Compared with oil prices well in excess of 100 dollars at times such risk premia are negligible and can be safely ignored.

Our ultimate question of interest is whether there is a time-varying risk premium. Evidence for or against a large average risk premium in the oil futures market does not resolve this question. On the one hand, even a large and statistically significant risk premium can typically be accommodated by conventional structural oil market models, provided the risk premium is constant over time. On the other hand, even an average risk premium of zero does not rule out the presence of a time-varying risk premium that averages to zero. Whether the risk premium is time-varying (or equivalently whether the excess returns are predictable) based on information available at time *t* typically has been assessed based on the regression

$$\left(F_t^h - S_{t+h}\right)/ S_{t+h} = \alpha + \beta x_t + v_{t+h}, \tag{1}$$

where the regressor x_t denotes a vector containing the set of candidate predictors that are conjectured to be correlated with excess returns. These predictors may reflect the state of the economy, of the oil market, or of related commodity markets. Much of the debate in the literature centers on the appropriate choice of x_t.[3]

There are three classes of models that have been used to estimate time-varying risk premia. The first class of models relies on the regression framework developed by Fama and French (1987, 1988) which has been adapted to oil futures markets by a number of subsequent studies. This framework focuses on the predictive power of the basis, defined as $\left(F_t^h - S_t\right)/ S_t$, for the futures returns, and does not require the user to specify any other predictors. The second class of models, which comprises much of the recent literature specifies return regressions of the form (1) involving a wider set of financial and macroeconomic predictors, some including the basis and some not. The third class of models relies on risk premium estimates from term structure models and does not require any return regressions.

[3] Typically, the objective in the risk premium literature has been to show that the predictive power of a variable of interest remains statistically significant even after the inclusion of other predictors in the regression, commonly referred to as *controls*. This outcome is interpreted as evidence of a causal effect from the predictor variable on the excess returns. It is clear, however, that in general predictive regression analysis is not suitable for answering questions of causality. Our analysis therefore abstracts from questions of causality and focuses on the predictive content of the proxies for oil-specific and systematic risk in oil futures markets proposed in the literature.

4.1. Regressing realized futures returns on the basis

Following Fama and French (1987, 1988) the traditional approach to testing for time-varying risk premia in commodity futures markets, including the crude oil market, has been based on one of the following two regression equations:

$$\left(F_t^h - S_{t+h}\right)/S_{t+h} = \alpha + \beta\left(F_t^h - S_t\right)/S_t + v_{t+h}. \tag{2}$$

$$\left(S_{t+h} - S_t\right)/S_t = \gamma + \delta\left(F_t^h - S_t\right)/S_t + u_{t+h}, \tag{3}$$

where the regressor $\left(F_t^h - S_t\right)/S_t$ is the basis. Equation (2) may be viewed as a special case of equation (1). Equation (2) is commonly referred to as the risk premium regression and equation (3) as the forecast efficiency regression. Evidence that $\beta > 0$ in the risk premium regression implies predictable variation in the realized excess returns and hence a time-varying risk premium. A constant average risk premium would be reflected in $\alpha \neq 0.$ The parameters in equations (2) and (3) are subject to an adding-up constraint. Because the sum of the excess returns and the change in the spot price equals the basis, it must be the case that $\beta + \delta = 1.$ Thus testing $H_0 : \delta = 1$ against $H_1 : \delta < 1$ in equation (3) is equivalent to testing $H_0 : \beta = 0$ against $H_1 : \beta > 0$ in equation (2).

Such tests have been reported in several studies of the oil futures market (e.g., Serletis 1991, Chernenko et al. 2004, Alquist and Kilian 2010). Table 2 shows that, after extending the estimation period to June 2014, one cannot reject $H_0 : \delta = 1$ against $H_1 : \delta < 1$ at any horizon. Although there is no evidence of a statistically significant time-varying risk premium at any horizon, the point estimates are consistent with some time variation in the risk premium.

4.2. Regressing futures returns on financial and macroeconomic predictors

The risk premium regression (2) in recent years has been increasingly replaced by regressions that allow for a variety of predictors of realized excess returns. These predictors include financial and macroeconomic variables that measure the state of the U.S. and global economy as well as oil-market specific variables that capture, for example, oil inventory dynamics, the degree of financialization, and hedging pressures in the oil futures market and in related markets. Table 3 summarizes the 15 monthly return regressions within this class of models on which our empirical analysis is based. Each model has been selected because it has been considered as supportive of the new consensus about a time-varying risk premium in oil markets in the literature.[4] The construction of the data is summarized in a not-for-publication appendix. As discussed in the introduction, we follow as closely as possible the definitions in the original studies, we update the data until June 2014, and we embed earlier studies within one coherent framework to obtain comparable estimates of the time-varying risk premium specific to the oil futures market.

4.3. Term structure models of the risk premium

A very different approach to estimating the time-varying risk premium in the oil futures market was proposed by Hamilton and Wu (2014a). Rather than specifying predictors of the excess return, they infer the risk premium indirectly from the observed time series properties of oil futures prices. Their premise is that some participants in this market use oil futures contracts to hedge oil price risk. The arbitrageurs who take the other side of these contracts receive compensation for their assumption of nondiversifiable risk in the form of positive expected returns from their positions. Building on Ang and Piazzesi (2003), Hamilton and Wu propose a model of this interaction that relies on an affine factor structure for oil futures prices. Time-varying risk premia are identified from differences between observed oil futures prices and the rational expectation of oil futures prices implied by the term structure model. In implementing

[4] Only two studies have proposed quarterly return models (see Etula 2013, Acharya et al. 2013). These studies are discussed in section 10.

this approach we rely on the same code as Hamilton and Wu (2014a), but we extend the sample to June 2014.

5. Alternative estimates of the time-varying risk premium

Although there is an emerging consensus in the literature that the risk premium in the oil futures market is time-varying, few studies report time series estimates of this risk premium. Often it remains unclear just how large the time-varying risk premium is because the authors are preoccupied with establishing the statistical significance of their preferred set of predictors and do not report a time series for the risk premium in U.S. dollars. Figure 2 allows a comparison of the estimates of the time-varying risk premium from the 16 monthly models discussed in sections 4.2 and 4.3 from January 1992 to June 2014. For expository purposes we focus on the 12-month horizon. Qualitatively similar results are obtained for all other horizons.

Figure 2 illustrates that, while there may be agreement on the presence of a time-varying risk premium in the oil futures market, there is substantial disagreement on the magnitude and sign of this time-varying risk premium. The disagreement is most pronounced after 2004. The standard deviation of the risk premium across alternative specifications ranges from \$0.52 in a given month to \$12.86. Alternative estimates of the risk premium may differ by as much as \$56 for the same month. Clearly, not all of these estimates can be equally valid. This raises the question of which estimates we can rely on and which ought to be discarded. Answering this question is essential for constructing a reliable measure of oil price expectations.

6. How to select the most accurate estimate of the risk premium

Many economic models require measures of how the market expectation of the price of oil, denoted by $E_t(S_{t+h})$, has evolved over time in the past. In the absence of a fully articulated economic decision problem, the conventional metric in assessing the accuracy of oil price expectations measures is their MSPE, defined as $E[S_{t+h} - E_t(S_{t+h})]^2$.[5] Standard arbitrage arguments imply that in the absence of a risk premium the oil futures price is the conditional expectation of the price of oil, $F_t^h = E_t[S_{t+h}]$ (e.g., Serletis 1991; Alquist and Kilian 2010). It is a well-known statistical result that the conditional expectation minimizes the MSPE under quadratic loss (see Granger 1969).[6] Hence, in the absence of a risk premium, F_t^h minimizes the MSPE.

This result breaks down in the presence of a time-varying risk premium, but, given that $F_t^h = E_t[S_{t+h}] + RP_t^h$ in this case, where RP_t^h denotes the time-varying risk premium in dollars, the risk-adjusted futures price $F_t^h - RP_t^h$ instead attains the minimum MSPE. While RP_t^h is not observable, it can be estimated, as discussed in section 4. Adjusting F_t^h for the estimated risk premium should systematically lower the MSPE compared with the conventional measure of oil price expectations, F_t^h. If $F_t^h - \widehat{RP}_t^h$, where $\widehat{RP}_t^h$ represents the full-sample estimate of the risk premium, does not have lower MSPE than F_t^h, the estimate of the time-varying risk premium and the expectations measure implied by this estimate clearly are not credible. Thus, biased or imprecise estimates of the time-varying risk premium would be expected to reveal themselves based on a comparison of the MSPEs of F_t^h and $F_t^h - \widehat{RP}_t^h$. This observation allows us to further discriminate between the many alternative estimates of the risk premium presented in section 5 and to resolve the uncertainty about how to choose between competing models of the time-varying risk premium. If a given risk premium estimate does not improve the accuracy of the

[5] Because this measure of oil price expectations is intended for use in modelling economic behavior and not in trading in real time, there is no reason to evaluate its accuracy based on alternative metrics such as the profits that might be realized by trading on this information.
[6] This result in fact applies more generally under any prediction error loss function that is symmetric about zero. For further discussion and a formal statement of this result also see Granger and Newbold (1986).

expectations measure at all horizons, this estimate is discarded from further consideration. Moreover, it is straightforward to determine the most plausible measure of the market expectation of the price of oil (and of the time-varying risk premium) among the remaining risk premium models by selecting the model that delivers the largest MSPE reductions.

6.1. How our approach differs conceptually from earlier studies

To the best of our knowledge this implication of the analysis in Granger (1969) has not been utilized previously in this context. Many studies of the time-varying risk premium focus exclusively on in-sample evidence. Some studies also report simulated out-of-sample results based on rolling or recursive regression estimates for the excess returns. It is common in this literature to interpret the simulated out-of-sample evaluation of the empirical risk premium model based on recursive and rolling regression estimates as a robustness check for the in-sample results. It is important to emphasize that our approach in this section, although superficially similar in that we report MSPE ratios for risk-adjusted futures prices, differs from these earlier studies.

First, unlike most of the studies in the finance literature on the oil risk premium, we are not concerned with the predictive power of the excess return regression, but with the ability of the risk-adjusted futures price to predict the spot price of oil. The latter question is only rarely addressed in the existing literature. One example is Pagano and Pisani (2009). This distinction is important because the model, $i \in \{1,...,N\}$, that minimizes the MSPE of the return regression $[(F_t^h - S_{t+h})/S_{t+h} - \widehat{RP}_{i,t}^h]^2$, where $\widehat{RP}_{i,t}^h$ is the fitted value from return regression i (and is expressed as a percent change or, equivalently, as a fraction), clearly is not in general the same as the model that minimizes the MSPE of the forecast of the spot price of crude oil in dollars based on the risk-adjusted futures price, given by $[S_{t+h} - F_t^h / (1+\widehat{RP}_{i,t}^h)]^2$.[7]

Second, unlike in Pagano and Pisani (2009) our objective in this section is not to form expectations about the of price oil beyond the end of the available estimation sample. Rather our objective is to recover the market expectations that prevailed in the past. This objective requires the use of all the information contained in the sample. If the econometrician's objective is to recover an estimate of the market expectation that prevailed historically, clearly the most efficient approach is to use regression estimates based on the full sample.[8] Predictive success as defined and measured in this section does not necessarily imply predictive success under conditions faced by applied forecasters, given the differences in the information set. It is possible for a researcher to obtain accurate measures of time series of historical oil price expectations in oil futures markets without being able to forecast the price of oil out of sample, as illustrated in section 9.

Third, our approach to selecting the most accurate expectations measure helps control for data mining in fitting excess returns. Searching for the most accurate return predictors inevitably invites overfitting, as researchers individually or collectively mine the data for the most successful predictive model of excess returns (see Inoue and Kilian 2004). We deal with this concern in two ways. First, because our construction of the oil price expectations measure relies on a different loss function than the loss function used in fitting the excess return, we

[7] This point is related to, but distinct from the result in Clements and Hendry (1998) that the ranking of two predictive models may change depending on whether we evaluate the MSPE of the growth rate or the log level of the dependent variable.

[8] One might object that estimates of market expectations should only reflect information available to market participants in real time. This view would indeed be compelling, if market participants were extracting signals about the risk premium (and hence about oil price expectations) from the same data as the econometrician. This is not the case, however. Rather oil futures market participants know their own oil price expectations and their risk preferences and trade on this basis. They have no need to estimate the market expectation. Only the econometrician wishing to estimate the risk premium from the oil market data faces a signal extraction problem.

automatically penalize models of the time-varying risk premium that suffer from overfitting. This approach is not unlike an out-of-sample evaluation, but does not require out-of-sample analysis. Second, we also assess the economic plausibility of the resulting expectations measure.

6.2. Empirical results

We use two measures of predictive success for S_{t+h}. One measure is the ratio of the MSPE of the risk-adjusted futures prices to the MSPE of the no-change forecast. This normalization is without loss of generality and facilitates the implementation of statistical tests for improvements in accuracy. The no-change forecast is the standard benchmark in studies of oil price forecasts, so it makes sense to follow this convention here (see Alquist et al. 2013). We assess the statistical significance of the MSPE reductions based on the test of Clark and West (2007). The other measure of predictive success is the success ratio, which represents an estimate of the probability with which the forecast correctly anticipates the direction of change in the price of oil. Any success ratio greater than 0.5 is considered an improvement in directional accuracy in that tossing a fair coin would result in a 50% chance of success in predicting the direction of change. The statistical significance of the directional accuracy relative to this benchmark can be assessed based on the test of Pesaran and Timmermann (2009).

A useful starting point is Table 4 which focuses on the basis regression also employed by Fama and French (1987, 1988). It can be shown that the risk-adjusted futures forecast in this case is algebraically equivalent to the oil price forecast $S_t\left(1+\hat{\gamma}+\hat{\delta}\left(F_t^h-S_t\right)/S_t\right)$, where $\hat{\gamma}$ and $\hat{\delta}$ refer to the least-squares estimates of the standard forecast efficiency regression (3). Table 4 shows that the risk adjustment actually increases the MSPE ratios compared with F_t^h, casting doubt on this measure of the risk premium. Very similar results are obtained if we restrict δ to unity. This evidence suggests that neither specification provides a credible measure of the time-varying risk premium.

Results for the extended return regressions and for the term structure model of Hamilton and Wu (2014a) are reported in Table 5. The model specifications are labelled as shown in Table 3. Table 5 shows that, after adjusting for the estimated risk premium, the MSPE ratios decline systematically across all horizons in only seven of sixteen cases. These results help discriminate further between models that previously have been considered largely equivalent. For example, although Bessembinder (1992) considered both specifications B1 and B2 a good fit, our analysis suggests that only B2 is a legitimate measure of the risk premium. The closely related specification employed in Sadorsky (2002) does not systematically improve on F_t^h either, whereas the specification of Bessembinder and Chan (1992) appears more successful. Likewise, of the three specifications reported in De Roon et al. (2000), only two systematically improve on the unadjusted oil futures price. The DNV3 specification is marginally more accurate than the DNV1 specification. Moreover, of the two specifications based on Gorton et al. (2013) only GHR1 appears promising. All three specifications explored by Pagano and Pisani (2009) and the model based on Bessembinder and Seguin (1993) fail to yield systematic MSPE reductions. Finally, of the two specifications favored by Hong and Yogo (2012) only HY2 works.

We conclude that many of the models proposed in the literature can be eliminated from consideration. Among the five most successful excess return regressions, the GHR1 model yields the most accurate measure of expectations. The single most accurate predictor in Table 5 by far, however, is based on the Hamilton and Wu (2014a) term structure model. The risk-adjusted forecast based on the HW model yields reductions in the MSPE ratio by an additional 19 to 36 percentage points compared with the futures price, depending on the horizon. The improvements in accuracy are not limited to the MSPE. At the same time the directional accuracy, measured by the probability with which the forecast correctly anticipates the direction of change in the price of oil, increases by 9 to 13 percentage points compared with F_t^h, depending on the horizon, which again is more accurate than any of the return regressions. All improvements in accuracy are statistically significant at the 5% level.

7. An Encompassing Approach to Predicting the Risk Premium

Table 5 suggests that there are some combinations of predictors that help predict the excess returns and hence the spot price. The model specifications considered in the literature, of course, do not span the space of possible models. In the interest of isolating the most useful predictors, we now construct two additional return regressions. The first regression encompasses all 30 individual predictors considered so far. This approach allows for the fact that many of the predictors are mutually correlated. We are not interested in the predictive power of individual regressors at this point so much as in their joint predictive power. The second regression is based on the subset of predictors in the first regression that are statistically significant at the 10% level. This pre-test approach mirrors the approach taken in the existing literature. We report additional results for both regressions in Table 6.

Table 6 shows that the all-predictors return model is substantially more accurate than any of the return regressions considered in the previous literature. It has substantially lower MSPE ratios than any of the models in Tables 4 and 5 with the exception of the HW model. It also has higher directional accuracy than any other return regression. Although the HW model remains more accurate at all but one horizon, the overall pattern of the MSPE ratios is qualitatively similar. We conclude that there are substantial gains in accuracy from considering a larger set of predictors.

A possible concern is that the number of regressors in the excess return regression is large. Because we do not evaluate the fit of the excess return regression, but rather the fit of the risk-adjusted futures price we do not have to be concerned that the predictive success of the model based on all predictors is caused by overfitting. Nevertheless, a natural question to ask is whether pre-tests may further improve the accuracy of the expectations measure. Table 7 shows that of the 30 predictors only the basis in oil futures markets appears statistically significant at all horizons based on two-sided t-tests. The CRSP returns, unexpected changes in U.S. industrial production, and the cross-market hedging pressure from the silver futures market are significant at all but one horizon. In addition, the OECD composite leading indicator and the change in the default premium are statistically significant at short horizons only, and the change in the 3-month T-bill rate only at longer horizons.

Although the results after pre-testing remain more accurate than any of the return regressions in Tables 4 and 5, dropping statistically insignificant predictors systematically lowers the accuracy of the expectations measure at all horizons compared with the unrestricted model in Table 6. We therefore focus on the unrestricted return regression and on the HW model for the remainder of the paper.

8. Oil Price Expectations

A central question of interest is how the market's oil price expectations have evolved since 1992 and between 2003 and 2008 in particular. To conserve space, we focus on the two most credible measures of oil price expectations derived in section 7. There is agreement across methodologies that the time-varying risk premium tends to increase with the horizon, but the extent of the variation differs across specifications, especially after 2005. Figure 3 shows the risk-adjusted forecast based on the return regression including all predictors for horizons $h \in \{3, 6, 9, 12\}$. It also reports the unadjusted F_t^h for comparison. The gap between these lines is the risk premium. Our discussion focuses on horizon 12 to conserve space. The 12-month-ahead time-varying risk premium is negligible until late 1998. Between January 1999 and December 2005, the time-varying risk premium fluctuates between -\$20 and +\$16, with an average of -\$6. These fluctuations increase substantially starting in January 2006. Temporarily, the risk premium reaches a low of -\$51 and a high of \$51. Clearly, these fluctuations are large enough to matter. There are indications that these estimates are implausibly volatile, however. For example, the 12-month-ahead oil price expectation reaches \$117 in October 2006, but drops to \$55 by November 2007 and continues to oscillate for the rest of the sample. In fact, oil price expectations at times are more volatile than the oil futures price, which does not seem plausible. Moreover, oil price expectations reach their all-time high of \$129 in April 2011, long after the peak in the spot price. It seems difficult to reconcile this pattern with public perceptions of oil price expectations.

In contrast, the risk-adjusted futures price based on the HW model in Figure 4 not only is more accurate than those based on all predictors, as shown in Table 6, but also implies a smoother evolution of oil price expectations. Between January 2003 and March 2008, during the surge in the spot price of oil, the time-varying risk premium on average is -\$12, although there are months when it reaches -\$28 toward the end of the subsample. During this time period, oil price expectations consistently exceed the oil futures prices. Between April 2008 and September 2008 the 12-month risk premium rises sharply to \$24 on average, reaching a peak of \$37. In other words, longer-horizon market expectations during this period were much lower than the oil futures price. For the remainder of the sample it fluctuates between -\$23 and +\$17 with an average of only -\$2. Oil price expectations were noticeably smoother than the oil futures price. We conclude that the HW model not only is more accurate in predicting S_{t+h}, but also delivers an economically more plausible measure of oil price expectations than the all-predictor model. The time series of oil price expectations in Figure 4 provides a practically feasible alternative to existing measures of oil price expectations used for economic modelling.

What did the market expect? Figure 4 shows that the 12-month-ahead oil price expectations increased with few exceptions throughout the period from early 2003 to mid-2008, from $30 initially to a peak of $100 in June 2008. While there is no evidence that the market anticipated the collapse of oil prices in the second half of 2008, even when the spot price unexpectedly reached $134 in June of 2008, market participants did not expect the price of oil to remain at that level for another year. After 2009, price expectations stabilized near $90.

Figure 5 examines to what extent the time-varying risk premium helps explain the persistent forecast errors of the oil futures price between early 2003 and early 2008 that we highlighted in the introduction. For expository purposes we focus on four specific points in time, each selected to be 15 months apart from the previous observation (April 2003, July 2004, October 2005, January 2007). Figure 5 illustrates that accounting for the HW risk premium typically helps close the persistent gap between oil price expectations and realizations, although there still are some points in time when substantial forecast errors occur even after accounting for the risk premium.

Finally, a key question is to what extent the evidence of time-varying risk premia in Figure 4 calls into question the reliability of structural VAR models of the physical market for crude oil that have been used to quantify the importance of speculative pressures in oil markets. These models exploit identifying assumptions motivated by theoretical models of the joint determination of changes in oil inventories, the spot price of oil and the oil futures price. These theoretical models postulate either a zero risk premium or a constant risk premium. The concern is less that these identifying assumptions would fail in the presence of a time-varying risk premium than that the VAR representation of the model becomes time-varying, invalidating the use of fixed-parameter models.

If we had found the time variation in the risk premium to be negligible at all horizons, there would no concern. The fact that the risk premium fluctuates substantially toward the end of the estimation sample used in Kilian and Murphy (2014) and Kilian and Lee (2014) complicates the analysis. How good of an approximation the constant-coefficient VAR model provides in this case depends on the extent to which variation in the risk premium translates to variation in the coefficients of the VAR model. The answer to this question is not known at this point. One way of shedding light on the quantitative importance of changes in the risk premium for the physical market for crude oil would be to develop a fully articulated calibrated general equilibrium model of the oil market with a time-varying risk premium and to examine their implied VAR representation. Developing such a model is beyond the scope of the current paper.

Alternatively, it may seem that this question could be addressed more directly by simply fitting a time-varying coefficient VAR model using the same identifying assumptions as in the earlier studies and comparing the results to the original fixed-coefficient model. This approach is not feasible, however, because this model involves many more parameters than existing estimation techniques can handle and because appropriate statistical summary measures for sign-identified time-varying parameter VAR models have yet to be developed.

9. Implications for Real-Time Oil Price Forecasts

Another reason to be interested in time-varying estimates of the risk premium is that they are potentially useful for improving the out-of-sample accuracy of oil price forecasts based on oil futures prices. The latter forecasts are commonly used at central banks and international organizations, for example. Unlike the analysis so far, producing risk-adjusted forecasts requires the recursive estimation of the time-varying risk premium using only data available to the forecaster at the time the forecast is generated. Table 8 shows that the risk-adjusted forecasts based on recursive estimates of the HW model have higher MSPE than the unadjusted futures price at all horizons. Estimating the time-varying risk premium based on a rolling window instead to accommodate the presence of a structural break in January 2005 identified by Hamilton and Wu (2014a) yields very similar results.

While the return regression including all 30 predictors is too highly parameterized to be estimated in real time, we also evaluated the excess return models in Table 7 as well as the unrestricted basis model in recursive and rolling regression settings. None of these models systematically improves on the real-time accuracy of F_t^h, even ignoring the data mining aspect of searching for the most accurate model over all alternative specifications.[9] The most successful model among all return regressions considered is the PP3 model based on the OECD composite leading indicator. Even that model, however, fails to improve upon the accuracy of F_t^h, except for a marginal reduction in the MSPE ratio at horizon 3. The results based on rolling return regressions are systematically less accurate than the recursive estimates and hence are not reported. We conclude that risk adjustments of real-time oil price forecasts cannot be recommended. There are, of course, other approaches to forecasting oil prices that have been shown to have superior real-time forecast accuracy and can be implemented by central banks and other forecasters (e.g., Baumeister and Kilian 2014, Baumeister, Kilian and Lee 2014).

10. Sensitivity Analysis

[9] These results also fail to account for the delays in the availability of the data for some of the predictors and subsequent revisions of the data. They should be viewed as an upper bound on the real-time forecast accuracy of the return regressions.

Although most empirical studies of the time-varying risk premium in the oil futures market rely on monthly return predictors, the studies by Acharya et al. (2013) and Etula (2013) are based on quarterly data. As the not-for-publication appendix shows, the quarterly results are very similar to the monthly results. Only some model specifications improve on the forecast accuracy of the quarterly futures price, and these individual models in turn are dominated by a return regression including all quarterly predictors. The most accurate measure of quarterly oil price expectations overall is obtained from the Hamilton and Wu (2014a) term structure model that also performed best in the monthly setting. The evolution of the quarterly HW time-varying risk premium is similar to the monthly estimates and the evolution of the implied oil price expectation is more plausible than for the all-predictor return regression. Finally, none of the quarterly models improve on the real-time accuracy of the quarterly oil price forecasts based on oil futures prices.[10]

11. Concluding Remarks

A time-varying risk premium renders the futures price a poor measure of the market expectation of the price of the underlying asset. Estimates of the risk premium may be used to construct improved measures of this expectation. Although there is a growing consensus that the risk premium in oil futures markets is time-varying, little is known about the evolution of this risk premium over time. We provided for the first time a systematic comparison of alternative estimates of the time-varying risk premium in the oil futures market. Our analysis focused on the most influential and most widely cited studies in this literature. We observed that some of these

[10] Our analysis does not include the return regression in Singleton (2014) based on changes in index fund and managed money spread positions, which is based on data that are available only since 2006. This sample is too short to assess the predictive content of this model for oil price expectations with any degree of reliability, prompting us to exclude this regression from consideration. Moreover, Hamilton and Wu (2014b) already showed that Singleton's in-sample results are not robust to updating the sample by two years to December 2011. Our efforts to update their analysis further to June 2014 failed because some of the required raw data are not publicly available (also see Sanders and Irwin 2013).

studies do not speak directly to the question of the risk premium in oil markets and extended the original analysis accordingly. All studies were compared on an extended sample extending from January 1992 to June 2014 using a consistent methodology.

We showed that there is tremendous variability in the risk premium estimates across model specifications, creating uncertainty about the magnitude of this risk premium as well as the implied market expectation of the price of oil. We showed that this model uncertainty can be resolved based on the observation that the risk-adjusted futures price is the conditional expectation of the price of oil and hence the minimum MSPE predictor by construction. This fact allowed us to rank alternative specifications based on their MSPE and to identify the most accurate measure of the market's oil price expectations. Our analysis revealed little empirical support for estimates of the risk premium based on excess return regressions of the type popular in recent applied work on oil markets. In fact, many of these specifications proved inadmissible based on the implied oil price expectations. Our preferred estimate of the risk premium was based on a term structure model of the oil futures market developed by Hamilton and Wu (2014a). The expectations measure implied by this model also has higher directional accuracy than other specifications and is economically more plausible. We highlighted that this new measure of the market expectation of the spot price of oil is useful in modelling expectations across a wide range of economic models dealing with energy prices. We also discussed implications of our findings for the debate about speculative pressures in the physical market for crude oil, and we showed how correcting for the time-varying risk premium helps explain the persistent errors implied by the use of futures prices as oil price expectations between 2003 and 2008.

In extracting the market expectation of the price of oil from the futures price, it is essential to estimate the risk premium based on the full sample. This approach provides the most efficient estimate of the oil price expected by the market at each point in time in the past, which is the relevant expectations measure, for example, in estimating models of purchases of automobiles as in Busse, Knittel and Zettelmeyer (2013) or in Allcott and Wozny (2014). This approach also yields the relevant risk premium measure for the debate on speculation in oil futures markets. In contrast, if the objective is to improve the accuracy of out-of-sample forecasts of the price of oil by risk adjusting the oil futures price, real-time estimates of the risk premium are required. Such estimates may be constructed based on recursive or rolling regressions possibly subject to delays in the availability of the data and revisions of preliminary data. Not surprisingly, estimating the risk premium in real time is more challenging than estimating it using the full-sample information. We found that even the risk-adjusted forecast based on the Hamilton and Wu (2014a) term structure model is unable to improve on the accuracy of the unadjusted oil futures price. Similar results hold for all other model specifications in a real-time setting. We concluded that the accuracy of forecasts based on the oil futures price cannot be improved by adjusting the futures price by real-time estimates of the risk premium.

Expectations play a key role in a wide range of forward-looking economic models. While we chose to illustrate our procedure for recovering the market expectation of the price in the context of the oil futures market, the underlying methodology is general and can be applied to futures prices for foreign exchange, interest rates and many other commodities when there is disagreement between alternative models of the time-varying risk premium.

References

Acharya, V., L. Lochstoer, and T. Ramadorai (2013), "Limits to Arbitrage and Hedging: Evidence from Commodity Markets," *Journal of Financial Economics*, 109, 441-465.

Allcott, H., and N. Wozny (2014), “Gasoline Prices, Fuel Economy, and the Energy Paradox,” forthcoming: *Review of Economics and Statistics*.

Alquist, R., and Kilian, L. (2010), “What Do We Learn from the Price of Crude Oil Futures?” *Journal of Applied Econometrics*, 25, 539-573.

Alquist, R., Kilian, L., and R.J. Vigfusson (2013), “Forecasting the Price of Oil,” in: Elliott, G., and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2, Amsterdam: North-Holland, 427-507.

Ang, A., and M. Piazzesi (2003), “A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables,” *Journal of Monetary Economics*, 50, 745-787.

Baumeister, C., and L. Kilian (2014), “Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach,” forthcoming: *Journal of Business and Economic Statistics*.

Baumeister, C., L. Kilian and T.K. Lee (2014), “Are there Gains from Pooling Real-Time Oil Price Forecasts?”, forthcoming: *Energy Economics*.

Bessembinder, H. (1992), “Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets,” *Review of Financial Studies*, 5, 637-667.

Bessembinder, H., and K. Chan (1992), “Time-Varying Risk Premia and Forecastable Returns in Futures Markets,” *Journal of Financial Economics*, 32, 169-193.

Bessembinder, H., and P. Seguin (1993), “Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets,” *Journal of Financial and Quantitative Analysis*, 28, 21-39.

Busse, M., Knittel, C.R., and F. Zettelmeyer (2013), “Are Consumers Myopic? Evidence from New and Used Car Purchases,” *American Economic Review*, 103, 220-256.

Chernenko, S.V., Schwarz, K.B. and J.H. Wright (2004), "The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk," International Finance Discussion Paper No. 1025, Board of Governors of the Federal Reserve System.

Chong, J., and J. Miffre (2010), "Conditional Risk Premia and Correlations in Commodity Futures Markets," *Journal of Alternative Investments*, 12, 61-75.

Clark, T.E., and K.D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291-311.

Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press.

De Roon, F.A., T.E. Nijman, and C. Veld (2000), "Hedging Pressure Effects in Futures Markets," *Journal of Finance*, 55, 1437-1456.

Diebold, F.X., and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Etula, E. (2013), "Broker-Dealer Risk Appetite and Commodity Returns," *Journal of Financial Econometrics*, 11, 486-521.

Fama, E.F., and K.R. French (1987), "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business*, 60, 55-73.

Fama, E.F., and K.R. French (1988), "Business Cycles and the Behavior of Metals Prices," *Journal of Finance*, 43, 1075-1093.

Fattouh, B., Mahadeva, L., and L. Kilian (2013), "The Role of Speculation in Oil Markets: What Have We Learned So Far?", *Energy Journal*, 34, 7-33.

Giacomini, R., and H. White (2006), "Tests for Conditional Predictive Ability," *Econometrica*, 74, 1545-1578.

Gorton, G., F. Hayashi, and G. Rouwenhorst (2013), "The Fundamentals of Commodity Futures Returns," *Review of Finance*, 17, 35-105.

Granger, C.W.J. (1969), "Prediction with a Generalized Cost of Error Function," *Operations Research Quarterly*, 20, 199-207.

Granger, C.W.J., and P. Newbold (1986), *Forecasting Economic Time Series*, 2nd ed., Academic Press: San Diego.

Hamilton, J.D., and J.C. Wu (2014a), "Risk Premia in Crude Oil Futures Prices," *Journal of International Money and Finance*, 42, 9-37.

Hamilton, J.D., and J.C. Wu (2014b), "Effects of Index-Fund Investing on Commodity Futures Prices," forthcoming: *International Economic Review*.

Hong, H., and M. Yogo (2012), "What Does Futures Market Interest Tell Us about the Macroeconomy and Asset Prices?" *Journal of Financial Economics*, 105, 473-490.

Inoue, A., and L. Kilian (2004), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", *Econometric Reviews*, 23, 371-402.

Kellogg, R. (2014), "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review*, 104, 1698-1734.

Kilian, L. (2014), "Comment on Francis X. Diebold's 'Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests'," forthcoming: *Journal of Business and Economic Statistics*.

Kilian, L., and Murphy, D.P. (2014), "The Role of Inventories and Speculative Trading in the Global Market for Crude Oil," *Journal of Applied Econometrics*, 29, 454-478.

Kilian, L., and T.K. Lee (2014), "Quantifying the Speculative Component in the Real Price of Oil: The Role of Global Oil Inventories," *Journal of International Money and Finance*, 42, 71-87.

Pagano, P., and M. Pisani (2009), "Risk Adjusted Forecasts of Oil Prices," *B.E. Journal of Macroeconomics*, 9, article 24, 1-25.

Pesaran, M.H., and A. Timmermann (2009), "Testing Dependence Among Serially Correlated Multicategory Variables," *Journal of the American Statistical Association*, 104, 325-337.

Piazzesi, M., and E.T. Swanson (2008), "Futures Prices as Risk-Adjusted Forecasts of Monetary Policy," *Journal of Monetary Economics*, 55, 677-691.

Sanders, D.R., and S.H. Irwin (2013), "Measuring Index Investment in Commodity Futures Markets," *Energy Journal*, 34, 105-127.

Sadorsky, P. (2002), "Time-Varying Risk Premiums in Petroleum Futures Prices," *Energy Economics*, 24, 539-556.

Serletis, A. (1991), "Rational Expectations, Risk and Efficiency in Energy Futures Markets," *Energy Economics*, 13, 111-115.

Singleton, K.J. (2014), "Investor Flows and the 2008 Boom/Bust in Oil Prices," *Management Science*, 60, 300-318.

Timmermann, A. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *Quarterly Journal of Economics*, 108, 1135-1145.

Table 1. Testing for an Average Risk Premium

	$\left(F_t^h - S_{t+h}\right)/S_{t+h} = \alpha + e_{t+h}$		$\sum_t \left(F_t^h - S_{t+h}\right)/T$
Horizon h	$\hat{\alpha}$	p-values for $H_0: \alpha = 0$	\$
3	-0.681	0.717	-0.73
6	-1.431	0.680	-1.60
9	-3.115	0.431	-2.55
12	-5.075	0.192	-3.51

NOTES: All test statistics have been computed based on HAC standard errors. The asymptotic p-values are for the two-sided t-test. All estimates are based on data for 1988.11-2014.6. Rejections of the null at the 10% level are highlighted in bold. For some horizons there are additional observations prior to November 1988. Including these observations does not affect the substance of the results.

Table 2. Testing for a Time-Varying Risk Premium Based on Forecast Efficiency Regressions

	$\left(S_{t+h} - S_t\right)/S_t = \gamma + \delta\left(F_t^h - S_t\right)/S_t + u_{t+h}$		
Horizon h	$\hat{\gamma}$	$\hat{\delta}$	p-values for $H_0: \delta = 1$
3	0.030	1.072	0.435
6	0.068	0.938	0.436
9	0.098	0.811	0.294
12	0.136	1.106	0.372

NOTES: See Table 1. The t-test of $H_0: \delta = 1$ is conducted against $H_1: \delta < 1$.

Table 3. Monthly Predictor Variables for Oil Futures Returns

Article	Model	Predictors
Bessembinder (1992)	B1	Returns on CRSP value-weighted equity index
	B2	Returns on CRSP value-weighted equity index Unexpected CPI inflation Change in expected CPI inflation Change in 3-month T-bill rate Change in the term structure (20YGB – 3-month T-bill) Change in default premium (Baa – 20YGB) Unexpected change in U.S. industrial production
Bessembinder and Chan (1992)	BC	Dividend yield on CRSP value-weighted equity index 3-month T-bill rate Junk bond premium (Baa – Aaa)
Sadorsky (2002)	S	Return on dividend yield on S&P 500 common stock portfolio Return on junk bond premium (Baa – Aaa) Return on 3-month T-bill rate Market portfolio excess return
De Roon, Nijman, and Veld (2000)	DNV1	Returns on S&P 500 stock price index Own-market hedging pressure Cross-market hedging pressure for gold, silver, platinum, heating oil
	DNV2	Own-market hedging pressure scaled by its standard deviation Own-market price pressure scaled by its standard deviation
	DNV3	DNV1 + own-market price pressure
Gorton, Hayashi, and Rouwenhorst (2013)	GHR1	Normalized U.S. crude oil commercial inventories (no SPR)
	GHR2	Own-market hedging pressure
Hong and Yogo (2012)	HY1	1-month T-bill rate Yield spread (Aaa – 1MTbill) Basis by horizon
	HY2	HY1 + growth rate of oil market dollar open interest
Pagano and Pisani (2009)	PP1	Degree of capacity utilization in U.S. manufacturing
	PP2	Term spreads (2YGB–1YGB, 5YGB–2YGB, 10YGB–5YGB)
	PP3	Composite leading indicator for OECD + 6 NMEs
Bessembinder and Seguin (1993)	BS	Ratio of trading volume of oil futures contracts to open interest by horizon

NOTES: The sample period is 1986.1-2014.6 except for the series from the CRSP database which are only available until 2013.12 and the series in BS which start only in 1989.9 for horizons 3, 6, and 9, and in 1992.4 for horizon 12.

Table 4. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Full-Sample Estimates of the Risk Premium
Evaluation Period: 1992.1-2014.6

Horizon h	No Risk Premium F_t^h	Time-Varying Risk Premium $S_t\left(1+\hat{\gamma}+\hat{\delta}\left(F_t^h-S_t\right)/S_t\right)$	Constant Risk Premium $S_t\left(1+\hat{\gamma}+\left(F_t^h-S_t\right)/S_t\right)$
		(a) MSPE Ratio	
3	**0.987**	1.037	1.034
6	**0.982**	1.074	1.075
9	**0.949**	1.068	1.074
12	**0.882***	1.038	1.036
		(b) Success Ratio	
3	**0.515**	**0.571**	**0.571**
6	**0.502**	**0.566**	**0.562**
9	**0.553**	**0.622****	**0.615**
12	**0.548**	**0.649***	**0.649***

NOTES: All MSPE ratios have been normalized relative to the monthly no-change forecast. Boldface indicates an improvement on the monthly no-change forecast. Statistically significant improvements test are marked using * (5% significance level) and ** (10% significance level). The underlying risk-premium estimates are based on the full sample. MSPE reductions are evaluated based on the tests of Diebold and Mariano (1995) and Clark-West (2007), as appropriate, and improvements in directional accuracy based on the test of Pesaran-Timmermann (2009).

Table 5. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Full-Sample Estimates of the Risk Premium
Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	B1	B2	BC	S	DNV1	DNV2	DNV3
				(a) MSPE Ratio				
3	**0.987**	**0.910***	**0.946**	**0.959****	**0.994**	**0.895****	**1.000**	**0.894****
6	**0.982**	**0.954***	**0.895****	**0.957***	**0.974****	**0.935****	1.004	**0.935****
9	**0.949**	**0.942***	**0.852***	**0.932***	**0.953***	**0.911***	**0.984****	**0.910***
12	**0.882***	**0.868***	**0.775***	**0.878***	**0.877***	**0.794***	**0.925***	**0.794***
				(b) Success Ratio				
3	**0.515**	**0.525**	**0.536**	**0.521**	**0.551**	**0.571****	**0.545**	**0.556**
6	**0.502**	**0.547**	**0.547**	**0.540**	0.498	**0.570**	**0.517**	**0.570**
9	**0.553**	**0.573**	**0.592****	**0.576**	**0.565**	**0.630***	**0.561**	**0.645***
12	**0.548**	**0.587**	**0.622***	**0.560**	**0.591**	**0.703***	**0.583**	**0.714***

Horizon h	GHR1	GHR2	HY1	HY2	PP1	PP2	PP3	BS	HW
				(a) MSPE Ratio					
3	**0.949****	1.003	**0.969***	**0.969***	**0.980****	1.002	**0.887***	**0.985**	**0.794***
6	**0.869***	1.011	**0.957***	**0.905***	**0.987**	**0.985**	**0.941***	**0.986**	**0.667***
9	**0.824***	**0.957****	**0.934***	**0.855***	**0.973**	**0.942****	**0.958****	**0.957****	**0.592***
12	**0.682***	**0.881***	**0.897***	**0.837***	**0.915***	**0.867***	**0.901***	**0.891***	**0.535***
				(b) Success Ratio					
3	**0.530**	**0.549**	**0.545**	**0.537**	**0.526**	**0.522**	**0.593***	**0.522**	**0.611***
6	**0.532**	**0.525**	**0.543**	**0.491**	**0.513**	**0.528**	**0.551**	**0.517**	**0.623***
9	**0.599***	**0.550**	**0.611***	**0.550**	**0.553**	**0.561**	**0.550**	**0.550**	**0.645***
12	**0.625***	**0.575****	**0.587**	**0.591**	**0.568**	**0.568**	**0.560**	**0.565**	**0.676***

NOTES: See Table 4. The evaluation period for B1, B2, BC and S for h=3 ends in 2013.12 and that for BS at h=12 starts only in 1992.4 due to data constraints.

Table 6. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Full-Sample Estimates of the Risk Premium
Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	All predictors	After pre-testing	HW
		(a) MSPE Ratio		
3	**0.987**	**0.777****	**0.804****	**0.794***
6	**0.982**	**0.705****	**0.792***	**0.667***
9	**0.949**	**0.670***	**0.874***	**0.592***
12	**0.882***	**0.612***	**0.652***	**0.535***
		(b) Success Ratio		
3	**0.515**	**0.619***	**0.566***	**0.611***
6	**0.502**	**0.634***	**0.585***	**0.623***
9	**0.553**	**0.634***	**0.622***	**0.645***
12	**0.548**	**0.718***	**0.710***	**0.676***

NOTES: See Table 4. The set of all predictors includes all variables in Table 3 except for market liquidity because the 12-month horizon liquidity series only starts in 1992.4. Including the latter variable with suitable changes in the evaluation period does not materially affect the results. The pre-test consists of a two-sided t-test at the 10% level. The results of the pre-test are reported in Table 7. The estimates for $h=3$ end in 2013.12, reflecting data constraints for some predictors.

Table 7. *p*-values for *t*-Tests on the Predictors in the Unrestricted Return Regression Model

		Horizon (Months)			
		3	6	9	12
B2	Returns on CRSP value-weighted equity index	**0.059**	**0.026**	0.191	**0.052**
	Change in 3-month T-bill rate	0.247	0.864	**0.085**	**0.004**
	Change in the term structure	0.133	0.637	0.854	0.719
	Change in default premium	**0.032**	**0.060**	0.491	0.303
	Unexpected change in U.S. industrial production	0.101	**0.042**	**0.018**	**0.019**
	Change in expected CPI inflation	0.743	0.206	0.585	**0.032**
	Unexpected CPI inflation	0.545	0.385	0.320	0.263
DNV1	Returns on S&P 500 stock price index	0.524	0.488	0.384	**0.025**
	Gold hedging pressure	**0.007**	0.114	0.778	0.646
	Silver hedging pressure	0.235	**0.007**	**0.034**	**0.007**
	Platinum hedging pressure	0.111	0.384	0.767	0.178
	Heating oil hedging pressure	0.668	0.416	0.734	0.638
	Own-market hedging pressure	0.952	0.491	0.446	**0.004**
GHR1	Normalized U.S. crude oil commercial inventories (no SPR)	**0.073**	0.166	0.114	**0.004**
HY1	1-month T-bill rate	0.926	0.660	0.476	0.547
	Yield spread	0.414	0.781	0.310	0.480
	Growth rate of oil market dollar open interest	0.246	**0.020**	**0.027**	0.171
	Basis	**0.041**	**0.041**	**0.012**	**0.008**
BC	Dividend yield on CRSP value-weighted equity index	0.656	0.681	0.684	0.219
	3-month T-bill rate	0.854	0.663	0.482	0.595
	Junk bond premium	**0.029**	0.731	0.840	0.321
S	Return on dividend yield on S&P 500 common stock portfolio	0.828	0.565	0.367	**0.023**
	Return on junk bond premium	0.377	0.259	0.551	**0.043**
	Market portfolio excess return	0.611	0.376	0.331	0.410
PP1	Degree of capacity utilization in U.S. manufacturing	0.888	0.422	0.189	**0.071**
PP2	Term spreads 1	0.517	0.735	0.488	0.333
	Term spreads 2	0.436	0.579	0.644	0.248
	Term spreads 3	0.671	0.477	0.325	0.438
PP3	Composite leading indicator for OECD + 6 NMEs	**0.002**	**0.009**	0.310	0.418
DNV3	Own-market price pressure	0.989	0.825	0.469	0.202

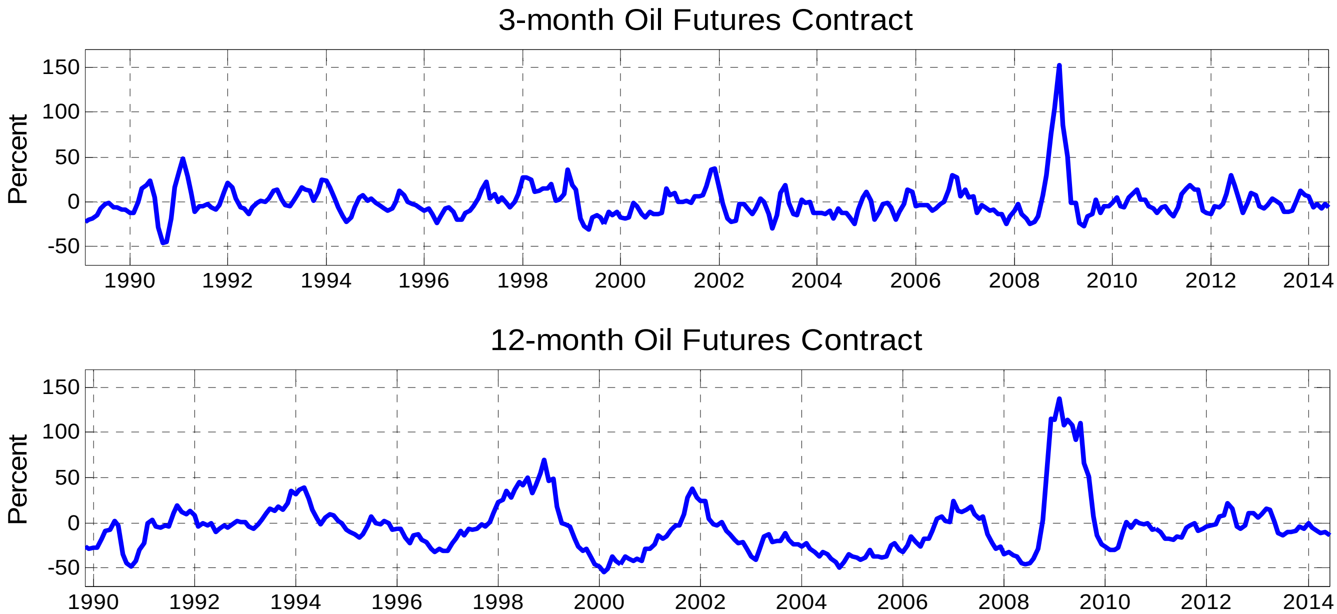
NOTES: The unrestricted return regression includes all 30 predictors included in the models in Tables 4 and 5 except for BS. Boldface indicates statistical significance of the two-sided t -test of the null of no predictability at the 10% level. Only the statistically significant predictors are retained in the return regression labeled "After Pre-testing" in Table 6.

Table 8. Risk-Adjusted Out-of-Sample Forecasts of the Spot Price of Oil: HW Model and Best Alternative Model Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	HW Recursive Window	HW Rolling Window (60 months)	PP3 Recursive Window
		(a) MSPE Ratio		
3	**0.987**	1.083	1.160	**0.951****
6	**0.982**	1.206	1.242	1.008
9	**0.949**	1.365	1.275	1.010
12	**0.882***	1.511	1.227	**0.955**
		(b) Success Ratio		
3	**0.515**	0.481	**0.504**	**0.575***
6	**0.502**	**0.502**	**0.547****	**0.532**
9	**0.553**	**0.573**	**0.588***	**0.542**
12	**0.548**	**0.556**	**0.579***	**0.541**

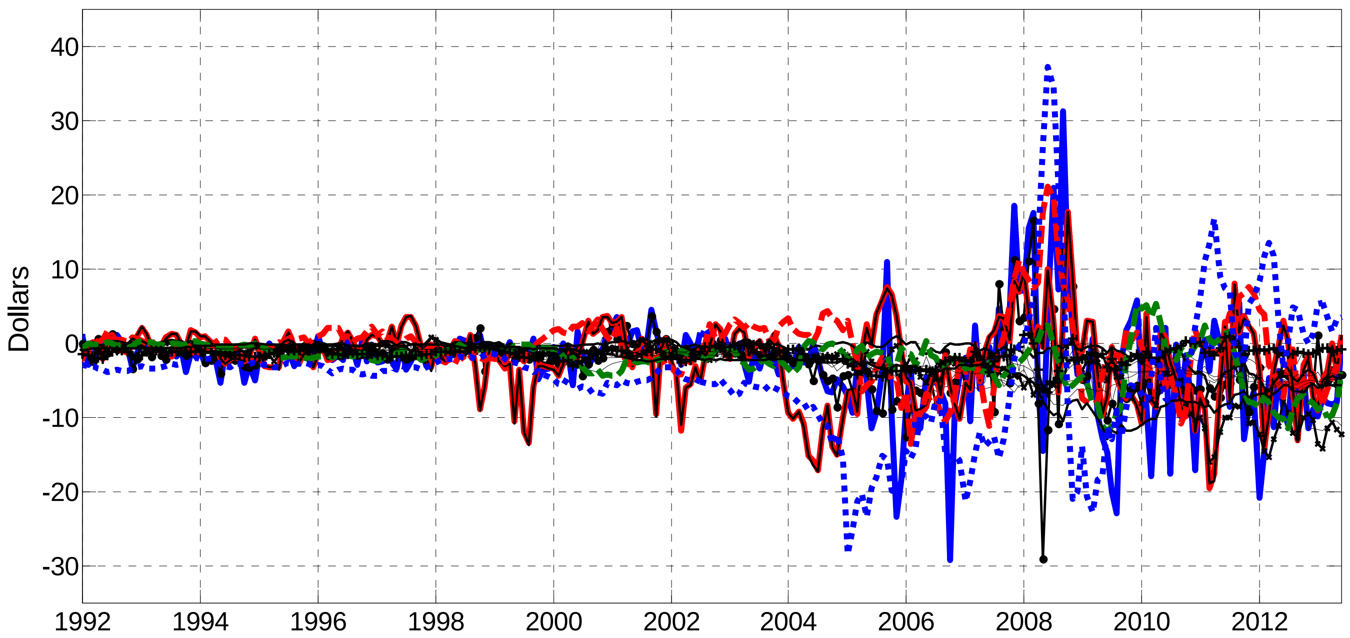
NOTES: Boldface indicates improvements on the no-change forecast. Statistically significant improvements test are marked using * (5% significance level) and ** (10% significance level). The initial estimation window ends in 1991.12. We follow the literature in assessing the statistical significance of the MSPE reductions based on the test of Clark and West (2007). This test (like similar tests in the literature) is biased toward rejecting the null of equal MSPEs because it tests the null of no predictability in population rather than the null of equal out-of-sample MSPEs (see Inoue and Kilian 2004). Thus, these test results have to be interpreted with caution. The alternative test of Giacomini and White (2006) does not apply either in our context because it does not allow for recursive estimation. For further discussion of the problem of out-of-sample inference see Kilian (2014). The directional accuracy of the forecasts is evaluated using the test of Pesaran and Timmermann (2009). PP3 is the most accurate model among all return regressions considered and one of only two such models with recursive MSPE ratios below 1 at more than one horizon. It is based on the OECD composite leading indicator. The results based on the rolling return regressions are much less accurate and hence are not reported.

Figure 1: Realized Excess Returns in the Oil Futures Market



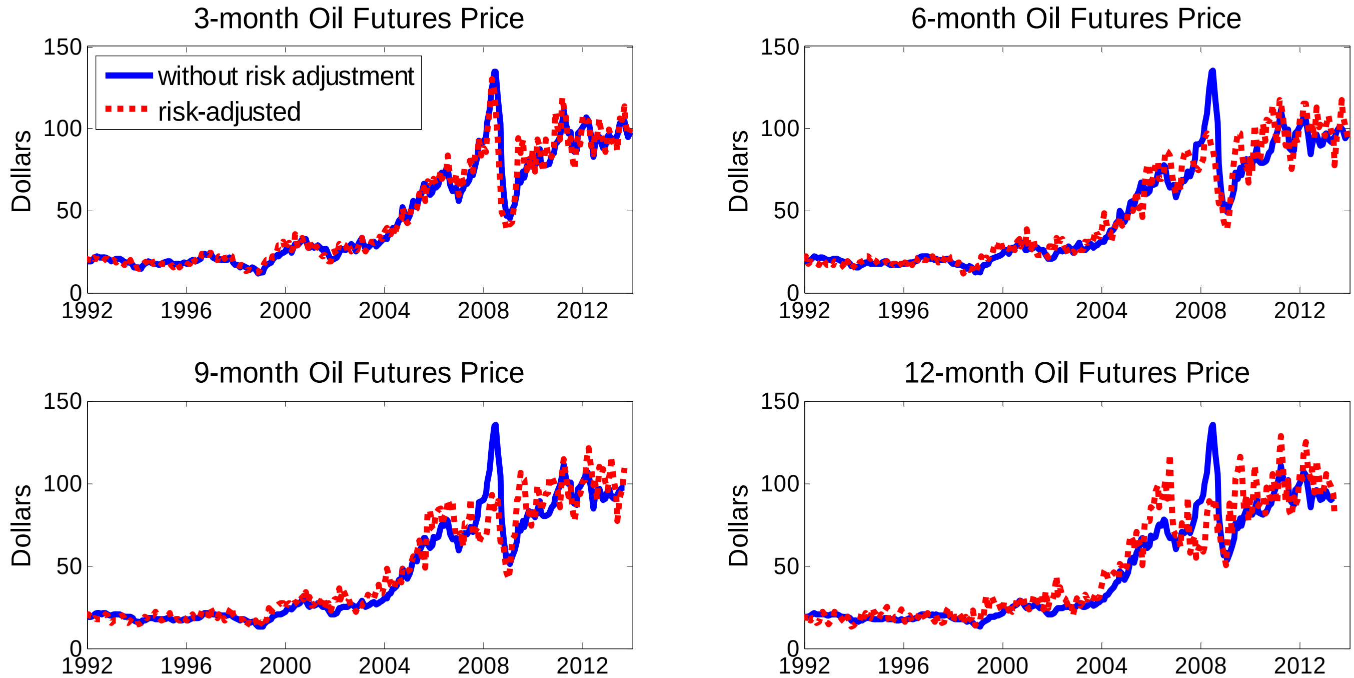
NOTES: The plot shows the realized excess returns on the 3-month and 12-month oil futures contract. The longer the horizon, the more persistent is the realized excess return.

Figure 2: Alternative Monthly Estimates of the Time-Varying Risk Premium in the Oil Futures Market at the 12-Month Horizon



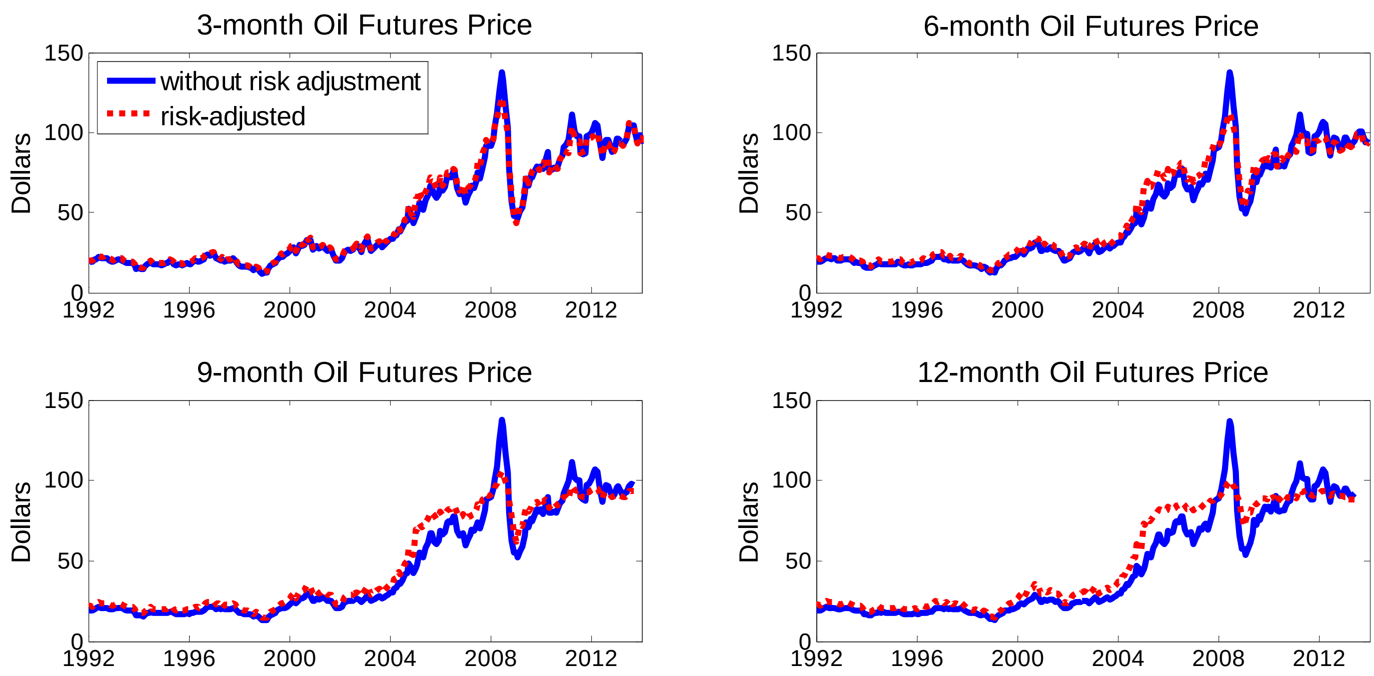
NOTES: The plot shows 16 alternative estimates of the time-varying risk premium that have been proposed in the literature. The plot illustrates that there is substantial disagreement on the magnitude and sign of the time-varying risk premium. Qualitatively similar results are obtained for other horizons.

Figure 3: Oil Price Expectations Based on the Return Regression with all Predictors



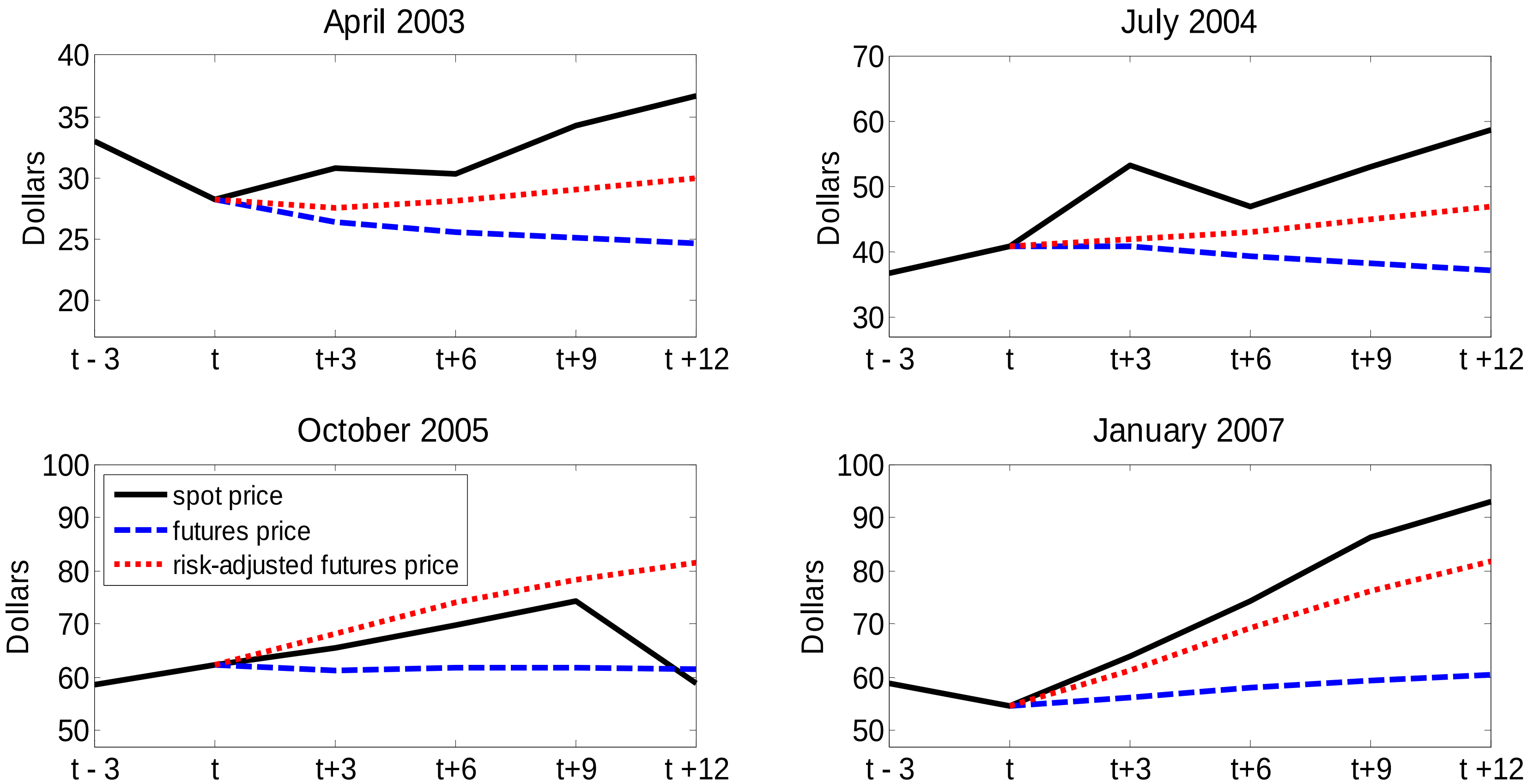
NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an unrestricted excess return regression including all 30 predictors listed in Table 7. These predictors were selected based on a review of the literature on the time-varying risk premium in oil markets.

Figure 4: Oil Price Expectations based on the HW Term Structure Model



NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an updated estimate of the term structure model in Hamilton and Wu (2014a).

Figure 5: Selected trajectories of F_t^h, the Realized Spot Price S_t, and $F_t^h - \widehat{RP}_t$, from HW Model



NOTES: The estimates of the time-varying risk premium are based on the term structure model of Hamilton and Wu (2014a) updated to June 2014. The risk-adjusted futures price is a measure of the market's oil price expectation at horizons of 3, 6, 9, and 12 months. The plot illustrates that adjusting the oil futures price for the HW estimate of the time-varying risk premium diminishes the systematic prediction errors that arise when using oil futures prices as proxies for oil price expectations.

Not-for-Publication Appendix

This appendix reports additional results for monthly and quarterly models to demonstrate the robustness of our main findings. It also discusses the data sources and the construction of the variables.

Table A.1. Risk-Adjusted Out-of-Sample Forecasts of the Nominal WTI Spot Price Based on Monthly Recursive Estimates of the Risk Premium Evaluation Period: 1992.1-2014.6

Horizon *h*	F_t^h	B1	B2	BC	S	DNV1	DNV2	DNV3
				(a) MSPE Ratio				
3	**0.987**	**0.985***	1.239	1.220	1.357	1.020	1.069	1.020
6	**0.982**	1.025	1.138	1.441	392.19	1.107	1.120	1.115
9	**0.949**	1.007	1.026	1.712	1.641	1.115	1.118	1.125
12	**0.882***	1.053	1.069	8.472	3.850	1.067	1.178	1.074
				(b) Success Ratio				
3	**0.515**	**0.502**	**0.540**	0.460	**0.528**	**0.541**	**0.575**	**0.534**
6	**0.502**	**0.525**	**0.551**	0.415	0.472	0.498	**0.547**	0.487
9	**0.553**	**0.553**	**0.573**	0.469	**0.519**	**0.557**	**0.580**	**0.550**
12	**0.548**	**0.556**	**0.571**	0.483	**0.529**	**0.564**	**0.625****	**0.591**

Horizon *h*	GHR1	GHR2	HY1	HY2	PP1	PP2	PP3	BS
				(a) MSPE Ratio				
3	1.074	1.055	1.073	1.098	1.069	1.125	**0.951****	-
6	1.131	1.150	1.147	1.139	1.129	1.175	1.008	-
9	1.106	1.196	1.179	1.195	1.141	1.306	1.010	-
12	1.144	1.011	1.200	1.564	1.167	1.477	**0.955**	-
				(b) Success Ratio				
3	**0.590***	**0.530**	0.470	0.463	0.470	**0.504**	**0.575***	-
6	**0.562**	0.483	0.460	0.464	0.479	**0.506**	**0.532**	-
9	**0.580**	**0.523**	0.496	**0.515**	**0.550**	**0.534**	**0.542**	-
12	**0.595**	0.498	0.382	0.452	**0.571**	**0.560**	**0.541**	-

NOTES: See Table 8. The evaluation period for B1, B2, BC and S at *h*=3 ends in 2013.12 due to data constraints. The initial estimation period for BS is too short for recursive estimation; therefore, this model is excluded. All variable transformations are performed only with the data available at each point in time. To obtain the normalized crude oil inventories a one-sided HP-filter with lambda=160,000 is used.

Table A.2. Quarterly Predictor Variables for Oil Futures Returns

Article	Model	Predictors
Etula (2013)	E1	Effective risk aversion (broker-dealer variable)
		S&P500 excess return (proxy for market risk)
	E2	Effective risk aversion (broker-dealer variable)
		Lagged oil futures returns
	E3	Effective risk aversion (broker-dealer variable)
		Lagged oil futures returns
		VIX implied volatility for S&P500
		3-month T-bill rate
		Yield spread (Aaa – 3-month T-Bill)
		Dividend yield on S&P 500 common stock portfolio
		U.S. CPI inflation
	E4	E2 + basis + own-market hedging pressure
Acharya, Lochstoer and Ramadorai (2013)	ALR1	Expected default frequency (EDF) in oil & gas industry
		Basis
		Default spread (Baa – Aaa)
		Median SPF forecast of quarterly GDP growth
		3-month T-bill rate
	ALR2	Zmijewski-score (Zm) of default risk in oil & gas industry
		Basis
		Default spread (Baa – Aaa)
		Median SPF forecast of quarterly GDP growth
		3-month T-bill rate
	ALR3	ALR1 + Realized quarterly variance of oil futures returns (RV) + Interaction between EDF and RV
	ALR4	ALR2 + Realized quarterly variance of oil futures returns (RV) + Interaction between Zm and RV
	ALR5	ALR1 + Effective risk aversion (broker-dealer variable) + Interaction between EDF and effective risk aversion
	ALR6	ALR2 + Effective risk aversion (broker-dealer variable) + Interaction between Zm and effective risk aversion

NOTES: The sample period is 1986.1-2014.2 except for the EDF series and the Zmijewski-score which are only available until 2009.4 and 2010.4, respectively.

Table A.3. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Quarterly Full-Sample Estimates
Evaluation Period: 1992.I-2014.II

Horizon h	F_t^h	HW	E1	E2	E3	E4
			(a) MSPE Ratio			
1	**0.984****	**0.822***	1.034	1.019	**0.910**	**0.901**
2	**0.983**	**0.685***	1.016	**0.988****	**0.896**	**0.901**
3	**0.954****	**0.587***	**0.972****	**0.930****	**0.789****	**0.755****
4	**0.905***	**0.526***	**0.933***	**0.893***	**0.644***	**0.675****
			(b) Success Ratio			
1	**0.589****	**0.600**	**0.633***	**0.656***	**0.611***	**0.544**
2	**0.517**	**0.618**	**0.573**	**0.584**	**0.607****	**0.573**
3	**0.545**	**0.682***	**0.614****	**0.580**	**0.602****	**0.568**
4	**0.540**	**0.724***	**0.609**	**0.586**	**0.701***	**0.690***

Horizon h	ALR1	ALR2	ALR3	ALR4	ALR5	ALR6	All predictors
				(a) MSPE Ratio			
1	**0.790**	**0.775**	**0.719**	**0.728**	**0.727**	**0.737**	**0.693**
2	**0.785**	**0.812**	**0.715**	**0.755**	**0.691**	**0.750****	**0.662**
3	**0.719**	**0.771****	**0.661**	**0.741**	**0.608**	**0.695****	**0.510****
4	**0.658****	**0.650***	**0.610***	**0.632****	**0.617****	**0.643***	**0.483****
				(b) Success Ratio			
1	**0.562**	**0.558**	**0.534**	**0.571**	**0.644***	**0.571**	**0.575**
2	**0.548**	**0.597****	**0.507**	**0.623****	**0.575**	**0.649***	**0.589**
3	**0.562**	**0.649***	**0.507**	**0.623****	**0.548**	**0.623***	**0.644***
4	**0.630**	**0.714***	**0.644**	**0.727***	**0.658****	**0.701***	**0.712***

NOTES: See Table 4. The evaluation period for ALR1, ALR3, and ALR5 ends h periods after 2009.IV and for ALR2, ALR4, and ALR6 ends h periods after 2010.IV due to data constraints. Adjusting the evaluation period underlying the MSPE ratios for F_t^h accordingly produces results qualitatively similar to the results shown in Table A.3.

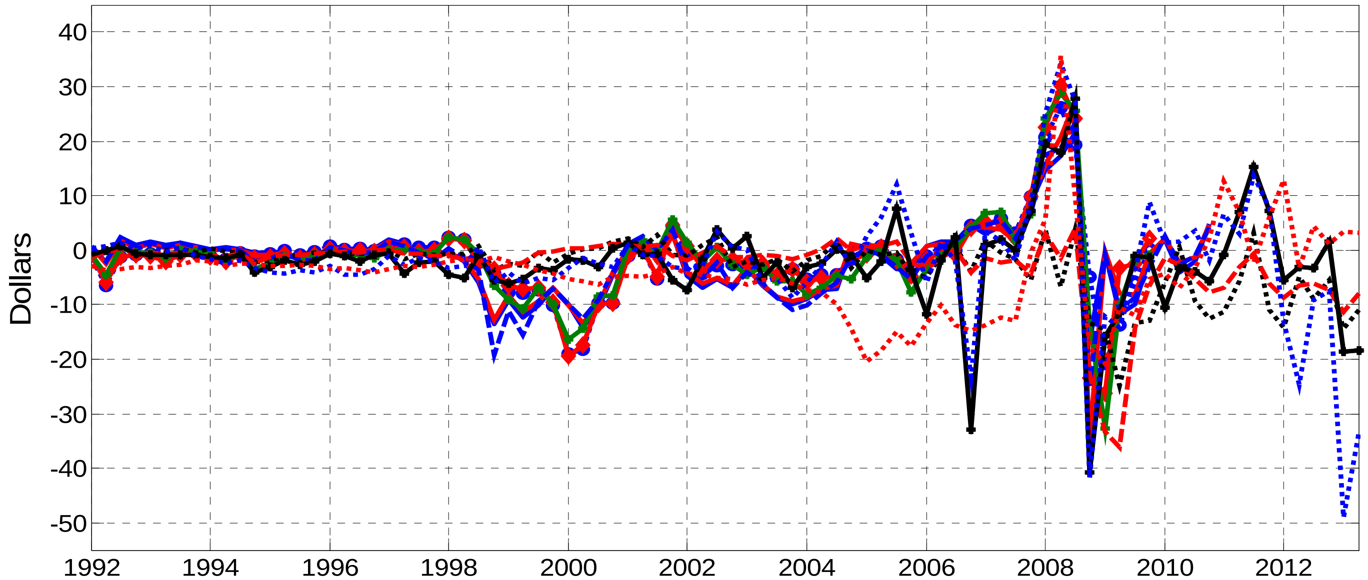
Table A.4. Risk-Adjusted Out-of-Sample Forecasts of the Nominal WTI Spot Price Based on Quarterly Recursive Estimates Evaluation Period: 1992.I-2014.II

Horizon h	F_t^h	HW	E1	E2	E3	E4
			(c) MSPE Ratio			
1	**0.984****	1.118	1.154	1.168	1.180	1.283
2	**0.983**	1.184	1.331	1.388	1.479	1.487
3	**0.954****	1.322	1.135	1.351	1.472	1.641
4	**0.905***	1.488	**0.995***	1.355	1.290	1.445
			(d) Success Ratio			
1	**0.589****	0.411	**0.544**	**0.544**	**0.556****	0.467
2	**0.517**	**0.528**	**0.573**	**0.551**	**0.506**	**0.539**
3	**0.545**	**0.568**	**0.602**	**0.591**	**0.534**	**0.534**
4	**0.540**	**0.598**	0.483	0.494	0.414	0.471

Horizon h	ALR1	ALR2	ALR3	ALR4	ALR5	ALR6
			(c) MSPE Ratio			
1	1.158	1.128	1.182	1.126	1.148	1.148
2	1.139	1.128	1.139	1.108	1.247	1.268
3	1.151	1.103	1.141	1.074	1.287	1.118
4	1.066	**0.944***	1.158	**0.996***	1.102	**0.920***
			(d) Success Ratio			
1	0.466	**0.558**	0.479	**0.571**	0.479	0.442
2	0.452	0.481	0.493	0.468	**0.534**	0.481
3	0.384	0.416	0.411	0.416	0.466	0.481
4	0.370	0.468	0.452	**0.519**	0.397	0.468

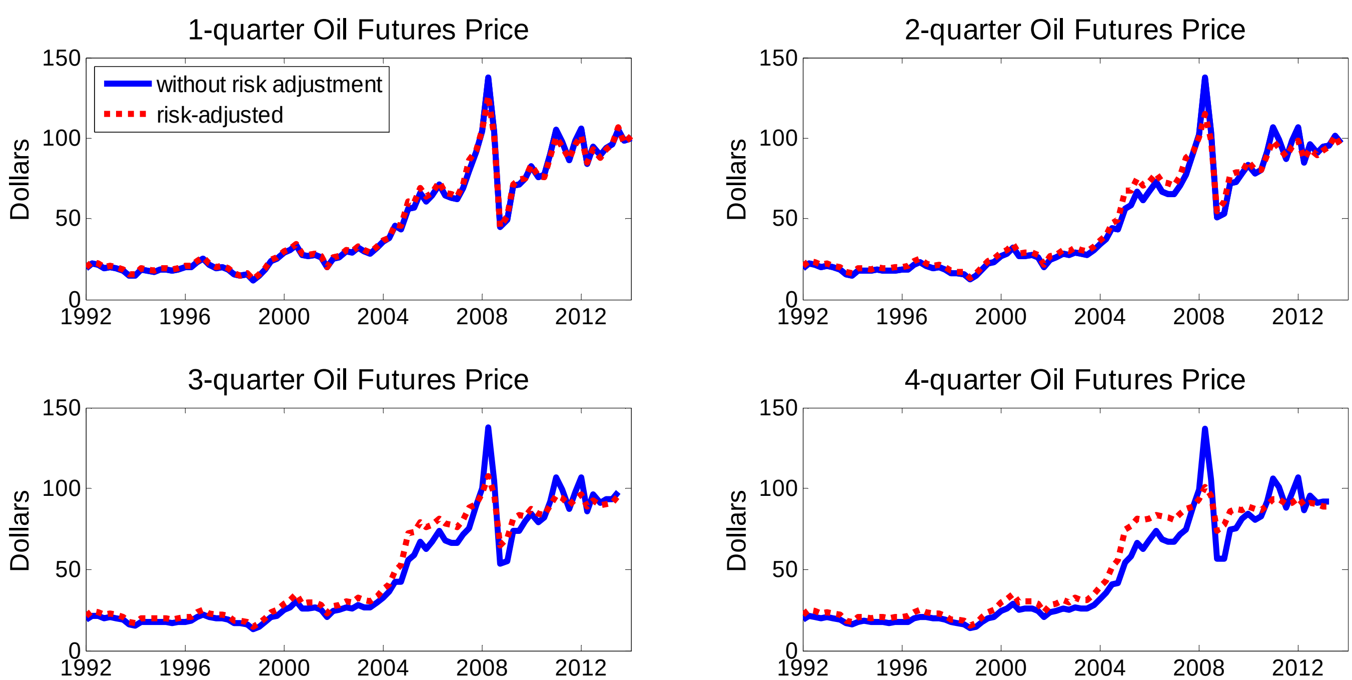
NOTES: See Table 8. All MSPE ratios have been normalized relative to the monthly no-change forecast. The evaluation period for ALR1, ALR3, and ALR5 ends h periods after 2009.IV and for ALR2, ALR4, and ALR6 ends h periods after in 2010.IV due to data constraints. Adjusting the evaluation period underlying the MSPE ratios for F_t^h accordingly produces results qualitatively similar to the results shown in Table A.4.

Figure A.1: Alternative Quarterly Estimates of the Time-Varying Risk Premium in the Oil Futures Market at the 4-Quarter Horizon



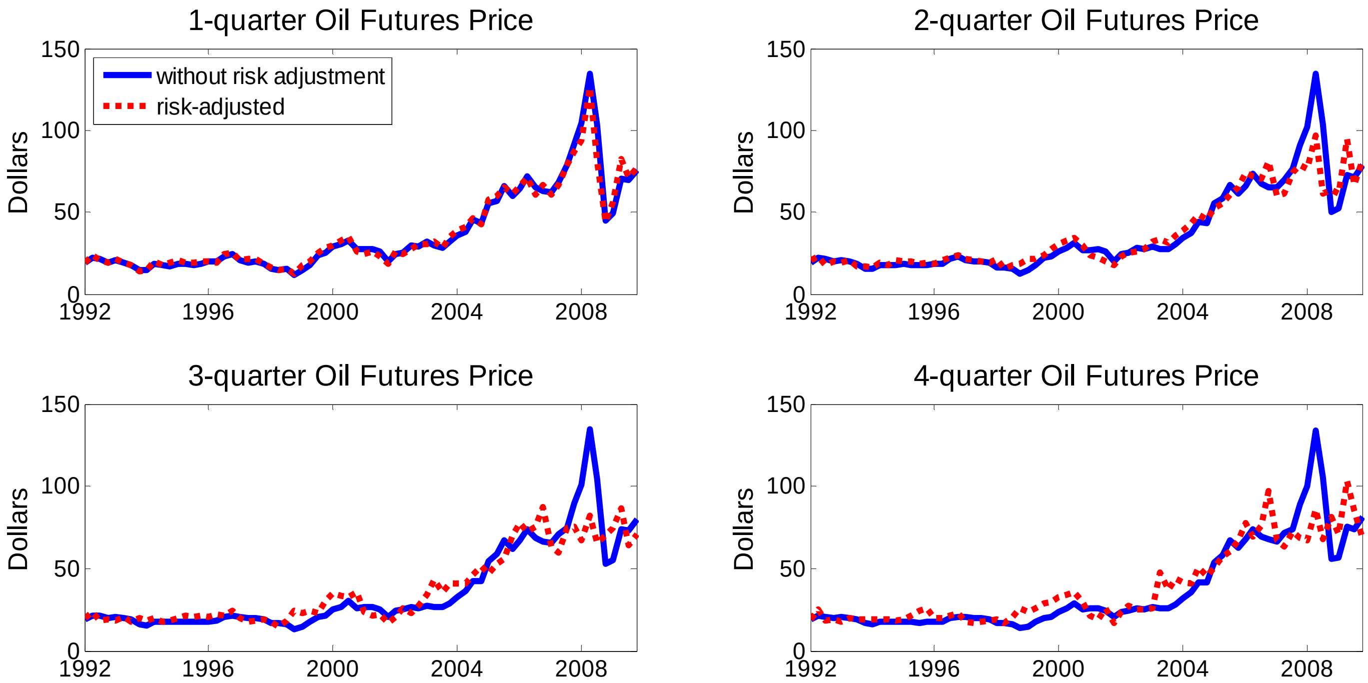
NOTES: The plot shows the 11 alternative estimates of the quarterly time-varying risk premium proposed in the literature, as summarized in Tables A.2 and A.3. The plot illustrates that there is substantial disagreement on the magnitude and sign of the time-varying risk premium. Qualitatively similar results are obtained for other horizons.

Figure A.2: Quarterly Oil Price Expectations based on the HW Term Structure Model



NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an updated estimate of the term structure model in Hamilton and Wu (2014a).

Figure A.3: Quarterly Oil Price Expectations based on the Return Regression with all Quarterly Predictors



NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an unrestricted excess return regression including all 19 predictors listed in Table A.2.

Data Sources and Construction of the Variables

Monthly averages of the daily WTI spot price were obtained from the U.S. Energy Information Administration (EIA) starting in 1986.1. The corresponding daily oil futures prices for maturities between 1 and 12 months are from Bloomberg going back as far as April 1983 depending on the maturity.

A. Macroeconomic and financial data

Monthly data for the 1-month T-bill rate, the 3-month T-bill rate, constant maturity rates for Treasury government bonds with maturities 1, 2, 5, 10 and 20 years, and Moody's Aaa and Baa corporate bond yields were obtained from the St. Louis Fed FRED database. From the same database we also retrieved data on the degree of capacity utilization in U.S. manufacturing measured in percent, U.S. consumer prices for all urban consumers and U.S. industrial production. Following Bessembinder (1992), annualized U.S. CPI inflation is decomposed into an expected and an unexpected component. Expected inflation is defined as the difference between the 1-month T-bill rate and the expected real interest rate which is determined as the simple average of ex-post realized real Treasury-bill returns over the preceding 12 months. Unexpected inflation is the difference between actual and expected inflation. The unexpected change in U.S. industrial production is computed based on the residuals from an ARIMA(3,1,0) model fitted to the raw industrial production data following Bessembinder (1992). The composite leading indicator for the OECD economies and six non-OECD economies (Brazil, China, India, Indonesia, Russia and South Africa) was obtained from the OECD *Main Economic Indicator* database. We follow Pagano and Pisani (2009) in selecting the ratio-to-trend series which expresses the cyclical behavior of the indicator as deviations from its long-run trend.

Returns on the CRSP value-weighted equity index with and without dividends are available via the Wharton Research Data Services. The dividend yield is constructed based on those two return series as described in Fama and French (1988). The market portfolio excess return used in Sadorsky (2002) is the equally weighted return on the CRSP value-weighted equity index (excluding dividends) and the Dow Jones-AIG Commodity Index in excess of the 3-month T-bill rate. The Dow Jones-AIG Commodity Index is composed of futures contracts on 19 physical commodities and is obtained from Bloomberg. Data for the S&P 500 stock price index and the dividend yield on the S&P 500 are from Haver Analytics. The VIX volatility index is obtained from Bloomberg.

The median 1-quarter-ahead forecast of U.S. real GDP growth from the *Survey of Professional Forecasters* is available from the Philadelphia Fed website. Etula's (2013) measure of effective risk aversion is based on balance sheet data for total financial assets and liabilities of U.S. security brokers and dealers as well as U.S. households and non-profit organizations and is obtained from the Federal Reserve's Flow of Funds database. The effective risk aversion index is constructed as: 1 + broker-dealer equity/household equity × (1 – broker-dealer leverage/market leverage), where equity is defined as the difference between financial assets and liabilities and leverage is computed as total assets over equity and the "market" refers to both U.S. security broker-dealers and households. A secular downward trend is removed by subtracting the 4-quarter moving average. The quarterly time series for the expected default frequency in the U.S. oil and gas industry and for the Zmijewski-score of default risk of U.S. oil and gas producers developed by Acharya, Lochstoer and Ramadorai (2013) are available on the authors' website (http://pages.stern.nyu.edu/~sternfin/vacharya/).

B. Commodity-specific data

Monthly data for U.S. crude oil commercial inventories excluding strategic petroleum reserves are reported in the *Monthly Energy Review* issued by the EIA. The normalized oil inventories are obtained by deseasonalizing and detrending the log of inventories using a Hodrick-Prescott filter with a smoothing parameter of 160,000 following Gorton, Hayashi and Rouwenhorst (2013).

Data on long and short positions of commercial traders for the crude oil, heating oil, gold, silver and platinum futures markets and open interest for the oil futures market are obtained from the reports on the *Commitment of Traders in Commodity Futures* published by the CFTC. Hedging pressure is defined as (number of short commercial positions – number of long commercial positions)/total number of commercial positions) for each market. Price pressure is the month-on-month change in hedging pressure. Following Hong and Yogo (2012) the oil market dollar open interest is computed as the total number of futures contracts outstanding times the spot price of crude oil; given that the monthly growth rate of this variable is noisy, they smooth it by taking the 12-month geometric average. Daily data on open interest and volume of trades of oil futures contracts for maturities 3, 6, 9 and 12 months are obtained from Bloomberg. The monthly liquidity measure is constructed as the median of the daily ratios of trading volume over open interest for each horizon. We follow Hong and Yogo (2012) in computing the basis as $(F_t^h / S_t)^{(1/h)} - 1$ for their return regression, whereas for the Fama and French regressions we define it as the percent spread between F_t^h and S_t without adjusting for the maturity, as in their original work.

CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

No.	Author(s)	Title
465	**Christopher D. Carroll, Jonathan A. Parker, Nicholas S. Souleles**	The Benefits of Panel Data in Consumer Expenditure Surveys
464	**Christopher D. Carroll**	Representing Consumption and Saving Without A Representative Consumer
463	**Piergiorgio Alessandri Sergio Masciantonio Andrea Zaghini**	Everything you always wanted to know about systemic importance (but were afraid to ask)
462	**Marti G. Subrahmanyam Dragon Yongjun Tang Sarah Qian Wang**	Credit Default Swaps and Corporate Cash Holdings
461	**Nils Friewald Rainer Jankowitsch Marti G. Subrahmanyam**	To Disclose or not to Disclose: Transparency and Liquidity in the Structured Product Market
460	**Lutz Kilian Robert J. Vigfusson**	The Role of Oil Price Shocks in Causing U.S. Recessions
459	**Simone M. Sepe Charles K. Whitehead**	Paying for Risk: Bankers, Compensation, and Competition
458	**Andrej Gill Matthias Heinz Heiner Schumacher**	Trust, Trustworthiness and Selection into the Financial Industry
457	**Laura Moretti**	Monetary Policy, Long Real Yields and the Financial Crisis

All CFS Working Papers are available at *www.ifk-cfs.de/publications--working-papers.html*.