University of Pretoria
Department of Economics Working Paper Series

Measuring Market Expectations

Christiane Baumeister
University of Notre Dame, University of Pretoria, NBER and CEPR

Working Paper: 2021-63
September 2021

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Measuring Market Expectations*

Christiane Baumeister
University of Notre Dame
University of Pretoria
NBER and CEPR

August 31, 2021

Abstract

Asset prices are a valuable source of information about financial market participants' expectations about key macroeconomic variables. However, the presence of time-varying risk premia requires an adjustment of market prices to obtain the market's rational assessment of future price and policy developments. This paper reviews empirical approaches for recovering market-based expectations. It starts by laying out the two canonical modeling frameworks that form the backbone for estimating risk premia and highlights the proliferation of risk pricing factors that result in a wide range of different asset-price-based expectation measures. It then describes a key methodological innovation to evaluate the empirical plausibility of risk premium estimates and to identify the most accurate market-based expectation measure. The usefulness of this general approach is illustrated for price expectations in the global oil market. Then, the paper provides an overview of the body of empirical evidence for monetary policy and inflation expectations with a special emphasis on market-specific characteristics that complicate the quest for the best possible market-based expectation measure. Finally, it discusses a number of economic applications where market expectations play a key role for evaluating economic models, guiding policy analysis, and deriving shock measures.

JEL classification: C52, E31, E43, E52, G14, Q43

Keywords: futures markets, risk premia, monetary policy, commodities, market expectations, financial markets, asset pricing, return regressions, affine term structure models, risk adjustment, model uncertainty, forecasting, expectational shocks

*I am grateful to Jim Hamilton for many insightful discussions and helpful suggestions. Christiane Baumeister, Department of Economics, 3060 Jenkins Nanovic Hall, Notre Dame, IN 46556, email: cjsbaumeister@gmail.com

1 Introduction

Expectations about the future play a key role in decision-making under uncertainty and form a building block for a wide range of forward-looking models in macroeconomics and finance. For example, expectations about future inflation will influence the price- and wage-setting behavior of firms and the consumption and storage decisions of households. Similarly, expectations about commodity price developments will affect production and investment plans as well as economic policy interventions. One valuable source of information about agents' expectations of future macroeconomic outcomes such as inflation, monetary policy, and commodity prices are assets traded in financial and futures markets. This paper surveys the literature on extracting market expectations from asset prices and summarizes existing empirical methods.

Section 2 reviews the asset pricing theory that forms the basis for the empirical analysis. In particular, it highlights the problem that financial instruments not only incorporate the rational expectation of market participants but also a compensation for undiversifiable risk. I revisit empirical efforts to uncover the presence of time-varying risk premia and provide an overview of modeling techniques used to separate out risk factors from the expectation component considering two classes of models for estimating time-varying risk premia, return regressions and Gaussian affine term structure models.

Section 3 addresses the question of which information set should be used for estimating the risk premium and introduces a general approach that allows to discriminate among different estimates. Using the oil futures market for illustration, I document the substantial disagreement on the magnitude and sign of the time-varying risk premium across predictor variables and discuss a methodology to identify the optimal estimate of the risk premium that allows to construct a unique and reliable measure of market expectations which can be used as input in economic decision problems. Specifically, I will show how standard forecasting tools can be applied to obtain the most credible estimate of the market price of risk and thus the implied market expectation. This statistical assessment can be supplemented by economic criteria to evaluate the plausibility of the implied market expectations.

Section 4 explores the universe of financial instruments that are useful for obtaining expectation measures. It provides a description of the characteristics of different financial and futures markets that directly influence the extraction of market expectations and discusses the difficulties that certain institutional features and trading frictions entail in accurately measuring expectations. In particular, I will focus on market-based monetary policy expectations and inflation expectations.

Section 5 presents a number of economic applications where market-based expectation measures have been used to address important policy questions. I start with a retrospective analysis of the oil market that ties the selected price expectation measure to historical events and changes in economic conditions to externally validate the usefulness of this market-based measure. I then consider a variety of contexts in which market expectations can be used as inputs for empirical exercises and theoretical models to test hypotheses, model decision problems, and inform policymakers. Specific

applications include the effects of regulatory policies on price expectations, such as the introduction of the Renewable Fuel Standard, the management of the U.S. Strategic Petroleum Reserve, and changes in gasoline taxes; the evaluation of the credibility of central bank policies based on the anchoring of agents' inflation expectations; the detection of speculative bubbles in commodity markets; the decisions related to household purchases of durable goods as well as capital investment and inventory accumulation; the derivation of surprise measures, such as monetary policy shocks and oil price expectation shocks, that capture revisions in market expectations and are key to understanding the transmission of shocks to the economy and designing policy responses; and the implications for real-time out-of-sample forecasts and macroeconomic projections generated by central banks and international organizations. In discussing these various applications, I will offer some suggestions for promising directions for future research.

Section 6 briefly concludes.

2 Market Expectations and the Price of Risk

It is common practice for central banks, international organizations, the private sector, and the financial press to treat the prices of futures, forwards, and other financial instruments as measures of market expectations. For example, Bernanke (2008) stresses that "policymakers and other analysts have often relied on quotes from commodity futures markets to derive forecasts of the prices of key commodities." This practice finds its origin in the notion that the price of a futures contract F_t^h with maturity h purchased at time t equals the expected value of the spot price S at expiry: $F_t^h = E_t(S_{t+h})$ where E_t denotes the expectations operator conditional on information available at time t. Thus, under the expectations hypothesis the expected payoff of holding a futures contract until maturity is zero since in efficient and rational financial markets, it is impossible to devise a trading strategy based on all relevant information that leads to making economic profits. This is equivalent to postulating that the prices of futures or forward contracts are unbiased predictors of future spot prices.

2.1 Testable Implications

This proposition can be tested using forecast efficiency regressions of the form:

$$\frac{(S_{t+h} - S_t)}{S_t} = \alpha + \beta \frac{(F_t^h - S_t)}{S_t} + \varepsilon_{t+h} \tag{1}$$

where the dependent variable is the realized percent change in the spot price between t and $t+h$, the independent variable is the current futures-spot spread expressed in percent changes and ε_{t+h} denotes the error term.[1] If futures prices are rational expectations of the future spot price, then we

[1] Much of the literature on forecast efficiency regressions uses log changes which are a good approximation if price changes are relatively small. However, some of the commodity prices considered in the empirical analysis below exhibit large fluctuations which makes the use of percent changes more appropriate.

would expect the joint hypothesis $H_0 : \alpha = 0,\ \beta = 1$ to hold. There are several other hypotheses of interest that can be tested based on these regressions. The null that the slope coefficient $\beta = 1$ implies that the futures price is an unbiased predictor of the future spot price, whereas the null that $\beta = 0$ implies that the futures-spot spread has no predictive content for future price changes. Deviations from the null of $\alpha = 0$ indicate that forecast errors contain a systematic component on average. The joint hypothesis $\alpha = 0$ and $\beta = 0$ means that any price changes are unpredictable.

We can test the validity of these hypotheses using commodity futures prices for a set of energy products (West Texas Intermediate (WTI) crude oil, Brent crude oil, gasoline, heating oil, and natural gas) and base metals (copper, lead, nickel, tin, and zinc) with maturities $h = 3, 6, 9$, and 12 months. I use futures prices on maturing contracts to measure spot prices as in Dusak (1973), Fama and French (1987), Hamilton (1992), and Chinn and Coibion (2014). Futures prices are sampled on the last trading day of month t and are obtained from Bloomberg. The start date differs across commodities and maturities in an effort to maximize the number of observations.[2] The sample ends in 2018.12 for all commodities.

The coefficients $\widehat{\alpha}$ and $\widehat{\beta}$ from estimating equation (1) by ordinary least squares are summarized in Table 1 for the energy commodities and in Table 2 for the base metals together with the p-values associated with the various hypotheses. Standard errors for all tests are computed based on the heteroskedasticity- and autocorrelation-consistent procedure of Newey and West (1987) with the lag truncation parameter set equal to the number of overlapping observations. We also report the R^2 to get a sense of the explanatory power of the futures-spot spread for the variation of future price changes.

For the crude oil market, the point estimates are positive and statistically significantly different from zero across all horizons for both WTI and Brent. While the slope coefficients are all less than 1, we cannot reject the null that the futures price is an unbiased predictor of the future spot price at any horizon. At the same time the futures-spot differential only explains a small fraction of subsequent price changes with the R^2 being at most 4% at short horizons and 7% at long horizons. We do reject the hypothesis of forecast efficiency for both crudes at all horizons. The joint hypothesis of no predictability is also decisively rejected. A similar picture emerges for heating oil except that at the 1-year horizon we cannot reject that the futures price is the rational expectation of the future spot price. For the gasoline market, the slope coefficients all exceed 1 and are significantly different from zero which indicates that the futures-spot spread has predictive power for future spot price changes. The spread accounts for a larger share of price variability with R^2's between 18% and 28%. The evidence for unbiasedness is somewhat less strong compared to the oil market and even rejected at the 9-month horizon. The null of market efficiency is strongly rejected. For the natural gas market, we can reject the hypothesis that β equals zero which suggests that there is useful information in the slope of the futures curve across all horizons. In fact, the quantitative ability of the futures-spot spread to explain future price changes is comparable to that

[2] Table 1A in the appendix provides details on the commodity- and horizon-specific start dates, the exchange where each commodity is traded, and the futures ticker used by Bloomberg.

of gasoline. We are not able to reject forecast efficiency and unbiasedness of natural gas futures prices except for $h = 6$.

Among the base metals, copper, lead, and tin yield negative slope coefficients which are statistically indistinguishable from zero, while the slope coefficients for zinc are well above 1. What is noteworthy is that even though we fail to reject the hypothesis that $\beta = 1$ for most metals except lead, the point estimates are actually quite far from one in most cases, suggesting that future price changes do vary with the slope of the futures curve. There is no statistical evidence of predictability of future price changes except for zinc where we can reject the joint hypothesis that $\alpha = 0$ and $\beta = 0$ at all horizons. While the null of forecast efficiency can be rejected for copper, lead, and tin at most horizons, futures prices for nickel and zinc can be considered rational expectations for future spot prices. The R^2 is extremely low for all base metals.

Taken together, the joint hypothesis of market efficiency is rejected for seven out of the ten commodities and there is overwhelming evidence for a predictable component in energy futures and zinc futures. The futures-spot price differential only accounts for a small proportion of subsequent price changes for most commodities, in particular for base metals.

This evidence is in line with previous findings in other contexts where the unbiased expectations hypothesis has also been consistently rejected. For evidence on forward interest rates refer to Fama and Bliss (1987), Froot (1989), Campbell and Shiller (1991), Chernenko, Schwarz and Wright (2004), Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), and Gürkaynak and Wright (2012), for federal funds futures to Sack (2004), Piazzesi and Swanson (2008), Ferrero and Nobili (2009), Hamilton (2009), and Hamilton and Okimoto (2011), for foreign exchange to Hansen and Hodrick (1980), Fama (1984), Korajczyk (1985), Froot and Frankel (1989), Bekaert and Hodrick (1993), and Chernenko et al. (2004), and for agricultural commodities and precious metals to Fama and French (1987, 1988), and Chinn and Coibion (2014), among many others.

2.2 Some Asset Pricing Basics

The theoretical justification for unbiased expectations is the presumption that risk-neutral market participants price any asset only based on the expected payoff without factoring in the uncertainty related to the randomness of the outcome and that any nonzero expected profits are arbitraged away. This assumption does not seem to hold in the data according to the results presented above.

One possible explanation is that financial market participants are risk averse.[3] Also in the case of risk aversion, the pricing of assets has to ensure the absence of profitable arbitrage opportunities which implies that there exists a strictly positive random variable M_{t+h} called the stochastic discount factor or pricing kernel that prices any asset at time $t < t + h$ with a stochastic payoff X_{t+h} at time $t + h$ accounting for variations in economic risk. For example, let $S_{t+h} - F_t^h$ be the random payoff of taking a long position in a forward or futures contract at time t for delivery h

[3] Chernenko et al. (2004) discuss several other potential explanations such as rational learning of market participants, irrational expectations, and "peso problems," but conclude that risk aversion is the most plausible one.

periods later and S_{t+h} the realized price at time $t+h$ for immediate delivery of the asset, then the equilibrium pricing condition is given by:

$$E_t(M_{t+h}(S_{t+h} - F_t^h)) = 0 \tag{2}$$

where E_t denotes the expectations operator conditional on time-t information.[4] Solving for the forward or futures price F_t^h results in:

$$F_t^h = \frac{E_t(M_{t+h}S_{t+h})}{E_t(M_{t+h})} = \frac{E_t(M_{t+h})E_t(S_{t+h}) + cov_t(M_{t+h}S_{t+h})}{E_t(M_{t+h})} = E_t(S_{t+h}) + \frac{cov_t(M_{t+h}S_{t+h})}{E_t(M_{t+h})} \tag{3}$$

which shows that under risk aversion the observed futures price equals the conditional expectation of the future spot price plus a risk premium that investors demand as compensation for taking on risky positions. This means that the evidence on predictability in future price changes and the rejection of forecast efficiency documented in Tables 1 and 2 is consistent with the existence of risk premia that arise from the exposure to non-diversifiable systematic risk factors. Time variation in risk premia can be explained by changes in both the risks and the investors' willingness to bear those risks over time.

A little detour. Söderlind and Svensson (1997, p. 398-399) provide a nice illustration of what determines whether the slope coefficient β in equation (1) is different from unity; in other words, to what extent the risk premium is time-varying. Assuming that the covariance is constant over time, they derive the following expression for the slope coefficient in the presence of a risk premium RP_t^h:

$$\beta = 1 - \frac{\sigma(\sigma+\rho)}{1+\sigma^2+2\rho\sigma} + \gamma$$

where

$$\sigma = \frac{std(RP_t^h)}{std\left(E_t\left(\frac{(S_{t+h}-S_t)}{S_t}\right)\right)} \qquad \text{and} \qquad \rho = corr\left(RP_t^h, E_t\left(\frac{(S_{t+h}-S_t)}{S_t}\right)\right)$$

and γ captures some systematic expectations errors, *i.e.* deviations of market expectations from rational expectations which can arise from small samples, learning, or truly irrational expectations but are generally considered small (see also Ferrero and Nobili, 2009). The extent to which β deviates from unity depends on the ratio of the standard deviation of the risk premium to the standard deviation of the expected spot price change (σ) and the correlation between the risk premium and the expected price change (ρ). Thus, a regression coefficient of one could result either from a constant risk premium ($\sigma = 0$) or from an offsetting combination of correlation and relative volatility ($\rho = -\sigma$) making the futures-spot spread an unbiased predictor. Figure 1 showcases how the regression coefficient adjusted for expectational errors ($\beta - \gamma$) evolves as a function of σ for different choices of ρ. A zero or positive correlation coefficient implies a monotonically declining value for the slope coefficient between 1 and 0. In this case, the estimated coefficient will be close

[4] If an asset has a fixed payoff at the time of maturity, e.g. an n-period zero-coupon bond, then the bond price today will only be a function of the pricing kernel: $P_t^n = E_t(\Pi_{i=1}^n M_{t+1}) \equiv E_t(M_{t+n})$; see Hamilton and Okimoto (2011) and Gürkaynak and Wright (2012) for more details and for a general treatment of bond prices.

to unity if the volatility of the risk premium is small relative to the volatility of the expected price change. Estimates of slope coefficients that are either negative or exceed one, as was the case for some of the base metals, are an indication that the correlation is negative. If the correlation is strongly negative, then small changes in σ around 1 can lead to large changes in the estimated β. For σ slightly below one, it is then possible to obtain coefficient estimates of 1. This provides some intuition for how the risk premium affects the slope coefficient β.

The previous discussion also highlights that testing for market efficiency or unbiasedness only makes sense in conjunction with specifying an asset pricing model which in the case of equation (3) amounts to postulating an economic model for the risk premium.[5] The decomposition in equation (3) makes it clear that if we had a direct measure of the risk premium or knew the process generating it, we could adjust the futures price and use it to infer the market's expected price. In fact, Dai and Singleton (2002) show that projections of risk-adjusted returns on the slope of the term structure of interest rates yields a β coefficient of unity, but only if the asset pricing model accurately captures the dynamic behavior of risk premia. This begs the question of how best to model risk premia.

2.3 Modeling Risk Premia

One of the key questions in empirical finance is what determines the risk premium and how to estimate it. There are two broad classes of models that have been developed for this purpose. The first class is return regressions which relate the ex-post risk premium to a set of observable factors that capture the investors' risk tolerance. The second class is dynamic term structure models that ensure the absence of arbitrage by imposing cross-equation restrictions and allow us to obtain an estimate of the risk premium based on the structural parameters of the model. I introduce both model frameworks, provide an illustration of their basic features, and show how they are related.

2.3.1 Return Regressions

This approach models the realized payoff or excess return as a linear function of possibly multiple, observable proxies for risk factors. For example, the final payoff on a long position in an h-period futures contract entered into at time t at rate F_t^h can be measured at the time of expiry $t+h$ as $(S_{t+h} - F_t^h)$ with the spot price S_{t+h} again being represented by the front-month contract F_{t+h-1}^1 and changes expressed in percentage terms:

$$\frac{F_{t+h-1}^1 - F_t^h}{F_t^h} = a_h + \sum_{k=1}^{K} b_{k,h}\ X_{k,t} + \epsilon_{t+h} \tag{4}$$

where a_h and b_h are horizon-specific regression coefficients, $\mathbf{X}_t$ is a vector containing up to K predictor variables available to market participants at time t, and ϵ_{t+h} is a mean-zero prediction error. The fitted value from this regression is an estimate of the time-varying risk premium.

[5] See Timmermann and Granger (2004) and Fama (2014) for a discussion of the joint hypothesis problem.

In theory, in the absence of trading frictions or liquidity factors, there should exist a single stochastic discount factor M_{t+h} that characterizes the risk premium in every financial asset as in (2). The theoretical value of M_{t+h} is a function of a postulated state vector that determines everything that could possibly happen in all aspects of the economy. In practice, for a particular class of assets, researchers focus on a small subset of variables thought to be most important in that market. Several risk factors considered are common across asset classes, while others are asset-specific and vary across markets and financial instruments.

A standard common factor derived from the capital asset pricing model (CAPM) is the overall return of the market portfolio, which is often measured by the CRSP value-weighted equity index (Bessembinder, 1992) or the S&P 500 stock price index (De Roon, Nijman and Veld, 2000). Another common factor that is often used is a measure of the overall level of economic activity (see, e.g., Campbell and Cochrane, 1999; Ferson and Harvey, 1991). Popular proxies for cyclical risk factors are employment growth (Piazzesi and Swanson, 2008), unexpected changes in industrial production (Bessembinder, 1992), the degree of capacity utilization in manufacturing (Pagano and Pisani, 2009), and indicators of U.S. real activity and global economic conditions (Hong and Yogo, 2012; Pagano and Pisani, 2009; Casiraghi and Miccoli, 2019; Baumeister, Korobilis and Lee, 2020), among others. Unexpected inflation is also considered a source of economic risk to the extent that inflation has real effects (Bessembinder, 1992). Financial indicators of the business cycle such as interest rates and yield spreads are also often used (Hong and Yogo, 2012; Pagano and Pisani, 2009; Piazzesi and Swanson, 2008), along with corporate bond spreads that measure changes in the risk of default in the economy (Bessembinder and Chan, 1992; Piazzesi and Swanson, 2008) and measures of stock market volatility that proxy for economic uncertainty (Casiraghi and Miccoli, 2019).

In commodity futures markets, one of the most widely used predictor variables is the slope of the futures curve (Fama and French, 1987). In bond markets, the analogous measure is the spread between long-term and short-term yields (Campbell and Shiller, 1991).[6] Asset-specific risks often arise from the types of traders who participate in a market since they have different characteristics and preferences. For example, commodity producers (e.g. farmers, oil producers, miners) and primary buyers (e.g. the food industry, airline companies, steel producers) want to hedge their physical exposure to the commodity by seeking insurance against price risk. A commodity's hedging pressure is determined by the relative size of positions taken in the futures markets by traders with a commercial interest. The most common empirical measure is the ratio of net short positions of commercial traders relative to all open positions. Two related risk factors are cross-hedging pressure which captures the influence of trading in closely related markets, and price pressure which results from any temporary changes in the demand or supply of futures contracts and can be approximated in the simplest case by changes in hedging pressure (De Roon et al., 2000). Such market imbalances can be caused by limits to financial arbitrage due to working capital constraints of broker-dealer firms (Adrian and Shin, 2010; Etula, 2013) and/or the risk

[6] Return regressions with the slope of the futures or yield curve as the only predictor are often treated as a separate class of models referred to as "basis regressions." I will use that terminology in Section 3.

appetite or hedging propensity of commodity producers (Acharya, Lochstoer and Ramadorai, 2013). The pertinent empirical proxies for the risk-bearing capacity of speculators and of producers are constructed based on balance-sheet data and market data. For financial futures, net long positions of *non*commercial market participants have been considered a relevant risk pricing factor (Piazzesi and Swanson, 2008). Overall trading activity measured as the amount of futures contracts outstanding times the spot price ("dollar open interest") is another potential determinant of the risk premium (Hong and Yogo, 2012). The trading strategies of new market participants or investor clienteles (e.g. exchange-traded funds, hedge funds, pension funds) might affect pricing dynamics in such a way as to change the equilibrium outcome. For example, the notional value of positions held by index-fund investors could influence the pricing of risk in commodity markets (Hamilton and Wu, 2013). Another market-specific friction for which investors demand compensation is liquidity risk. Proxies for market liquidity include relative transaction volume (Gürkaynak, Sack and Wright, 2010; Pflueger and Viceira, 2016), the ratio of the volume of futures contracts traded to open interest (Bessembinder and Seguin, 1993; Chinn and Coibion, 2014), and asset-swap spreads to infer financing costs (Pflueger and Viceira, 2016), among others. A fundamental determinant of risk for financial instruments that reference storable commodities such as oil, gasoline, industrial metals, and agricultural products, are inventory levels and dynamics motivated by the convenience yield implicit in the theory of storage (Acharya et al., 2013; Dincerler, Khokher and Simin, 2020; Gorton, Hayashi and Rouwenhorst, 2013; Pindyck, 2001).

This provides an illustration of the type of risk factors that have been proposed in the literature. This list is by no means exhaustive since new risk factors and their corresponding empirical measures are added all the time.[7] These risk factors are not all equally successful at predicting returns. I will explore the usefulness of a subset of these risk factors for modeling risk premia in oil futures markets in Section 3.

2.3.2 Gaussian Affine Term Structure Models

The idea underlying this class of models is that a small set of m latent or observed factors x_t jointly determine all asset prices in the economy in an internally-consistent way whose dynamics can be described as a Gaussian first-order vector autoregression (VAR):

$$x_{t+1} = c + \rho x_t + \Sigma u_{t+1} \qquad\qquad u_{t+1} \sim N(0, I_m) \tag{5}$$

Since all assets are priced in the same way, we can focus on a single asset class without loss of

[7] The proliferation of risk factors in the empirical asset pricing literature has been termed the "factor zoo" by Cochrane (2011) and is most prevalent in explaining the cross section of expected stock returns. For example, Harvey and Liu (2019) document the existence of over 400 potential factors published in top journals. There is a recent effort to systematically comb through the universe of candidate factors and select only the ones that are relevant for pricing risk based on their marginal importance using appropriate testing procedures (see Feng, Giglio and Xiu, 2020).

generality to illustrate the basic principles underlying this modeling framework.[8] For example, log oil futures prices $f_t^h \equiv \ln F_t^h$ for a contract entered into at time t that matures h periods later are assumed to be affine functions of these factors in the following way:

$$f_t^h = \alpha_h + \beta_h' x_t \tag{6}$$

with the factor loadings given by

$$\alpha_h = \alpha_{h-1} + \beta_{h-1}'(c - \lambda) + \frac{1}{2}\beta_{h-1}'\Sigma\Sigma'\beta_{h-1} \tag{7}$$

$$\beta_h' = \beta_{h-1}'(\rho - \Lambda) \tag{8}$$

where $\lambda_t = \lambda + \Lambda\, x_t$ are the market prices of risk that summarize the investor's attitudes toward risk and are also affine functions of the fundamental factors x_t that make them vary over time. These expressions are derived from the first-order condition of an investor who cares about the mean and variance of his portfolio.[9] Specifically, the notional investment amount $z_{h,t} \equiv F_t^h \cdot \#barrels$, which generates a cash flow of $(F_{t+1}^{h-1} - F_t^h)$ per barrel at time $t+1$, is chosen such as to maximize:

$$z_{h,t} E_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right] - (\gamma/2) z_{h,t}^2 Var_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right]$$

which implies the following first-order condition:

$$E_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right] = \gamma z_{h,t} Var_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right]. \tag{9}$$

Assuming that the log changes in oil futures prices are conditionally normally distributed based on the information set Ω_t, $(f_{t+1}^{h-1} - f_t^h)|\Omega_t \sim N(\mu_{h,t}, \sigma_{h,t}^2)$, Hamilton and Wu (2014, 2015) show that

$$E_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right] \approx \mu_{h,t} + \frac{1}{2}\sigma_{h,t}^2 = \alpha_{h-1} + \beta_{h-1}' x_{t+1} - \alpha_h + \beta_h' x_t + \frac{1}{2}\beta_{h-1}'\Sigma\Sigma'\beta_{h-1}$$

where the equality follows from the asset pricing equation (6) and the factor dynamics (5), and[10]

$$Var_t \left[\frac{F_{t+1}^{h-1} - F_t^h}{F_t^h} \right] \approx \sigma_{h,t}^2 = \beta_{h-1}'\Sigma\Sigma'\beta_{h-1}.$$

[8] The majority of papers in this literature study one asset or market at the time. Exceptions are, for example, Hamilton and Wu (2014) who consider an arbitrageur who takes positions in a range of different assets, and Christensen et al. (2010) and D'Amico et al. (2018) who jointly model investments in real and nominal bonds.

[9] This discussion as well as the notation closely follows the exposition in Hamilton and Wu (2014, 2015). For details on the empirical implementation and estimation that the results in Section 3 are based on, the reader is referred to Hamilton and Wu (2014). Their code is available at http://econweb.ucsd.edu/~jhamilto/hw4_code.zip

[10] The first expression is the expected value of a log-normally distributed variable F:

$E[F] = \exp[E(\ln F) + 1/2Var(\ln F)] \Longleftrightarrow \ln E[F] = E(\ln F) + 1/2var(\ln F) \Longleftrightarrow \ln E[\exp(f)] = E(f) + 1/2Var(f)$ with $f \equiv \ln(F)$.

Substituting these expressions into the first-order condition (9) and recognizing that Hamilton and Wu (2014) posit that in equilibrium the investor's risk exposure λ_t to positions taken in $z_{h,t}$ is also an affine function of the vector of factors as defined above yields:[11]

$$\alpha_{h-1} + \beta'_{h-1}x_{t+1} - \alpha_h + \beta'_h x_t + \frac{1}{2}\beta'_{h-1}\Sigma\Sigma'\beta_{h-1} = \beta'_{h-1}\lambda_t. \tag{10}$$

Replacing x_{t+1} with its expected value $c + \rho x_t$ from (5) and using $\lambda_t = \lambda + \Lambda\ x_t$ leads to the expressions for α_h and β'_h in equations (7) and (8). The recursions implied by these equations have to hold for every h and are the cornerstone of all affine term structure models since they allow to price all the assets in the economy ruling out arbitrage possibilities (see, e.g., Piazzesi, 2010; Gürkaynak and Wright, 2012). Specifically, the restriction that the cross-sectional factor loadings are functions of the parameters describing the state dynamics ensures dynamic consistency.

If investors were risk neutral, the same recursions would still apply and assets would still be priced according to (6) but under a modified law of motion for the factors

$$x_{t+1} = c^Q + \rho^Q x_t + \Sigma u^Q_{t+1} \qquad\qquad u^Q_{t+1} \sim N(0, I_m)$$

where the adjustments $c^Q = c - \lambda$ and $\rho^Q = \rho - \Lambda$ result from risk aversion. This means that in a risky environment investors behave as if the asset offers a lower expected payoff than it effectively does. What the pricing kernel M_{t+1} does is describe the mapping between the true data-generating process, also known as the P-measure, and the risk-neutral distribution, also known as the Q-measure (see Ang and Piazzesi, 2003).[12] Intuitively, the pricing kernel reweights the objective probabilities implied by the true distribution such that some outcomes have a higher probability than they objectively do which implies that investors require some compensation in that state of the world. Time-varying risk premia are then obtained as the difference between observed futures prices and the rational expectation of futures prices implied by the estimated term structure model when setting $\lambda = \Lambda = 0$.

2.3.3 An Integrative View

Hamilton and Wu (2015) use equation (10) to provide a unifying perspective by illustrating how the term structure framework relates to the return regression approach. Specifically, substituting equations (5) and (6) into (10), they show:

$$f^{h-1}_{t+1} - f^h_t = \kappa_{h-1} + \delta'_{h-1}x_t + \varepsilon^{h-1}_{t+1} \tag{11}$$

with $\kappa_{h-1} = \beta'_{h-1}\lambda - \frac{1}{2}\beta'_{h-1}\Sigma\Sigma'\beta_{h-1}$, $\delta'_{h-1} = \beta'_{h-1}\Lambda$, and $\varepsilon^{h-1}_{t+1} = \beta'_{h-1}\Sigma u_{t+1}$. Dai and Singleton (2002) draw a similar parallel for affine models of the yield curve in their equation (21), where

[11] Specifically, $\Sigma\Sigma'\beta_{h-1}\gamma z_{h,t} \equiv \lambda_t$ and the investment positions comove with the underlying factors.

[12] Most papers in this literature postulate a particular functional form for the pricing kernel. For more details, the reader is referred to Ang and Piazzesi (2003) and Gürkaynak and Wright (2010). For a more technical treatment of affine term structure models, see Piazzesi (2010).

they show that risk premia formulations resulting from Gaussian dynamic term structure models imply the same structure as the excess return regressions in Fama (1984) and Fama and Bliss (1987). A key difference between these two modeling frameworks is that affine term structure models impose cross-equation restrictions that rule out arbitrage strategies which means that at all horizons h the coefficients depend on the same set of structural parameters that describe the state dynamics and risk premia, while the coefficients in return regressions are obtained from unrestricted least-squares regressions estimated separately for each maturity h. Another difference is that return regressions rely on observed proxies for relevant risk factors, whereas in affine term structure models the unobserved factors are commonly inferred from the behavior of asset prices themselves. While often a small set of factors is sufficient to describe the term structure dynamics, many variants of empirical term structure models exist and have evolved over time to incorporate additional observable determinants beyond the information contained in the cross section of asset prices. For example, Ang and Piazzesi (2003) include factors extracted from a panel of inflation and real activity measures in addition to three latent yield curve factors (typically referred to as level, slope, and curvature). Instead, Bernanke, Reinhart and Sack (2004) only rely on a set of observed macroeconomic variables including GDP growth, inflation, the federal funds rate, and survey-based expectations of future inflation and growth, within a standard term structure model of interest rates.[13] Affine term structure models are flexible and can be tailored to account for different market characteristics (e.g. liquidity, segmentation) and specific institutional features (e.g. calendar irregularities) without having to resort to observed proxies. At the same time, the advantage of using observed variables to model risk premia is that we can pinpoint the source(s) of risk, whereas giving economic content to the latent factors in term structure models is often more difficult.[14]

While these modeling frameworks enable us to separate risk premia from market expectations, the previous discussion makes it clear that different models will produce different estimates of time-varying risk premia which imply different expectation measures. This raises the important question of how to choose among different measures of market expectations which I will turn to next.

[13] While this practice helps with the interpretability of the fundamental drivers in affine term structure models and establishes a link between asset price movements and macroeconomic dynamics (albeit often only unidirectional), it raises the question whether macroeconomic variables are truly factors for the purpose of yield curve modeling, often referred to as the "spanning hypothesis" (see, e.g., Rudebusch and Wu, 2008; Ludvigson and Ng, 2009; Bauer and Hamilton, 2018).

[14] An active strand of this literature concerned with macro-finance term structure models links the factor dynamics more explicitly to structural macroeconomic models and/or derives the pricing kernel from utility maximization which greatly enhances economic interpretability (see Gürkaynak and Wright (2012) for an overview of such models).

3 Extracting Measures of Market Expectations from Asset Prices

Using equation (4) which has the same basic structure as the generalized version of equation (11) that nests both modeling frameworks, we solve for the futures price at expiry to obtain

$$F^1_{t+h-1} = F^h_t(1 + a_h + \mathbf{b}'_h\mathbf{x}_t + \epsilon_{t+h})$$

where $\mathbf{x}_t$ refers to a $(K \times 1)$ vector of observable proxies or latent risk factors. Taking expectations on both sides and using the approximation $F^1_{t+h-1} \approx S_{t+h}$ yields the market's current expectation of the h-period-ahead spot price as the risk-premium-adjusted futures price:

$$E_t(S_{t+h}) = F^h_t(1 + a_h + \mathbf{b}'_h\mathbf{x}_t)$$

where the risk premium in dollars is $RP^h_t = F^h_t - E_t(S_{t+h})$. It is evident that there will be as many implied price expectation measures as there are risk premium estimates which can differ substantially across model specifications and which cannot all be equally valid. The key challenge is to select the most credible estimate of the market expectation for a given set of candidate risk premium models.

3.1 A General Approach to Identifying Market Expectations

Baumeister and Kilian (2017) propose a systematic approach to evaluating the plausibility of different estimates of time-varying risk premia drawing on insights from the forecasting literature. A conventional metric in assessing the accuracy of price expectations is their mean-squared prediction error (MSPE), defined as $E[S_{t+h} - E_t(S_{t+h})]^2$. The central idea is that the conditional expectation minimizes the MSPE under quadratic loss which is a well-known result in statistics (see, e.g., Granger, 1969; Granger and Newbold, 1986). This theoretical result allows us to rank alternative model specifications based on their MSPE and to resolve the model uncertainty that arises from the multiplicity of candidate risk factors. The most accurate measure of the implied market expectation will be the one that delivers the largest MSPE reduction.

It is important to note that this procedure does not involve generating out-of-sample forecasts. The implied market expectations are obtained based on the full-sample estimates of each risk premium model which provides the most efficient estimate of the price expected by the market at each point in time. While we are concerned with in-sample performance, the model rankings cannot be derived from the fit of the risk premium models because the proposed methodology uses a different loss function which makes it akin to an out-of-sample evaluation.

3.2 An Illustration Based on the Oil Market

We illustrate the usefulness of this approach for identifying the best possible, unique measure of the market's expectation about the future path of the spot price of crude oil along the lines of Baumeister and Kilian (2017), but for a larger universe of risk premium models and an updated sample

that includes the 2014-16 oil price decline. Consistent with the existing literature on estimating the oil risk premium, our analysis focuses on the WTI price of crude oil.[15] To obtain estimates of the time-varying risk premium for horizons $h = 3, 6, 9$, and 12 months, we rely on (*i*) basis regressions, (*ii*) payoff regressions for a selected set of the most prominent risk factors applied to commodity markets and oil in particular, which are summarized in Table 3, and (*iii*) the Hamilton and Wu (2014) term structure model for the oil futures market.[16] The evaluation period is 1997.8-2018.12.

Figure 2 shows the resulting estimates for the risk premium at the 12-month horizon across all 25 model specifications. Given our definition of the risk premium, a value of -\$3 indicates that the price expected by market participants was \$3 higher than the quoted futures price. While there is broad agreement across models that the risk premium was small, negative on average, and relatively stable, fluctuating moderately between -\$8 and \$1, before 2004, there is a notable increase in volatility thereafter. The most striking feature in the evolution of risk premia after 2005 is the substantial heterogeneity in magnitude and sign. Alternative estimates can differ as much as \$55 in a given month. The two episodes with the largest disagreement in risk compensation across model specifications are July 2008 and June 2014 where the estimated premia range from -\$28 to \$27 and from -\$10 to \$43, respectively. Mapping these risk premium estimates into market expectations by adjusting the futures price would result in an equally dispersed set of price expectations that cannot all be representative of the market's assessment of the future course of oil prices. Thus, Figure 2 provides a graphical illustration of the problem highlighted above that we are now going to resolve by evaluating the predictive success of the price expectations implied by the various models.

Table 4 presents the ratio of the MSPE of the model-implied price expectation to the MSPE of the monthly no-change prediction. This normalization is standard in forecasting and facilitates the implementation of statistical tests for improvements in accuracy (see, e.g., Baumeister et al., 2020). A ratio below 1 indicates that the model does better than a random walk, while a value above 1 indicates that it does worse. To gauge the statistical significance of reductions in MSPE, we use the tests of Diebold and Mariano (1995) and Clark and West (2007), as appropriate.

The first row of Table 4 reports the results for the unadjusted futures price which is the relevant benchmark against which we evaluate the price expectation estimates derived from the given set of risk premium models. The futures price outperforms the no-change prediction at each horizon with statistically significant gains in forecast accuracy that increase as the horizon lengthens. The two basis regressions only beat the no-change prediction at the 12-month horizon but do not imply competitive expectation measures since their MSPE ratios exceed those of the unadjusted futures

[15] We use monthly averages of WTI futures prices for consistency with the predictor variables in the payoff regressions whose reference period is the month. Data for variables available at a higher than monthly frequency are also averaged.

[16] Hamilton and Wu (2014) conduct their analysis on two subsamples with the first ending in December 2004 and the second starting in January 2005 to allow the structural parameters to change in response to the increased financialization of commodity markets after 2004. We impose the same break in estimation to account for the changes in oil futures price dynamics. The model is estimated at weekly frequency and the risk premium estimates are averaged over the month.

price. Among the payoff regressions, a larger set of predictor variables is useful at short horizons than at long horizons with about two-thirds of the models producing expectation estimates with lower MSPE ratios than the random walk at horizon 3 compared to only a handful at horizons 9 and 12. The most promising model specifications across all horizons include financial and macroeconomic indicators of the business cycle and measures of hedging pressure and trading activity. The implied price expectations yield statistically significant reductions in MSPE ratios by an additional 9 percentage points in the short run and 7 percentage points in the long run relative to the non-risk-adjusted futures price. The most successful model is the HW term structure model which attains the smallest MSPE ratios with impressive gains in accuracy of 17% at the 6-month horizon, 24% at the 9-month horizon, and 30% at the 12-month horizon compared to the no-change prediction, which are highly statistically significant. This amounts to improvements on the futures price between 14 and 16 percentage points. Only at horizon 3 is the HW model tied with two specifications proposed by Hong and Yogo (2012), while it outperforms them by a large margin at all other horizons. Based on our selection criterion, the oil price expectation implied by the HW model is the most credible measure of market expectations overall.

In addition to the statistical evidence, Baumeister and Kilian (2017) suggest to check the plausibility of the preferred market-based measure of oil price expectations based on economic criteria which might be particularly useful if the MSPE ratios are tied. They make the case that longer-term oil price expectations should not shift abruptly but rather evolve smoothly over time except in times of major market turmoil. More specifically, they should be less volatile than the unadjusted futures prices. This is indeed the case. The standard deviation of the expected price 1-year-ahead recovered from the HW model is 15% lower than that for the corresponding futures price. Instead, the price expectation measure that is ranked second according to the statistical criterion (HY7) is just as volatile as the futures price. Another useful check is to examine whether the expected price developments align with the historical narrative. I will provide a retrospective analysis of the evolution of market participants' oil price expectations in Section 5.1.

4 Existing Empirical Evidence for Selected Markets

There exists a long list of assets traded on financial, forward, and futures markets whose prices incorporate expectations about key macroeconomic variables such as inflation, house prices, freight costs, commodity prices, interest and exchange rates. The same general methodology for selecting the most plausible market-based expectation measure can be applied to the set of price expectation estimates derived for each variable of interest based on the broad model classes discussed in Section 2. Care must be exercised in accounting for specific features of each market in determining the relevant set of expectations for evaluation. To illustrate this point, I focus on monetary policy expectations and inflation expectations.

4.1 Monetary Policy Expectations

The futures market for federal funds has long been the primary source for gauging the future course of monetary policy given that the payoff at maturity is directly tied to the actual average fed funds rate, the policy instrument of the Federal Reserve, realized over the delivery month (see, e.g., Kuttner, 2001; Faust, Swanson and Wright, 2004; Sack, 2004; Piazzesi and Swanson, 2008; Hamilton and Okimoto, 2011). In contrast to oil and other commodities, futures contracts for federal funds are not the only financial instruments whose rates are influenced by traders' views about near-term changes in Fed policy and thus can be used to extract monetary policy expectations. There exists a range of alternative securities available for this purpose like Treasury bills, eurodollar deposits and futures, commercial paper, and term federal funds loans that differ in their credit quality and trading activity.[17] Thus, it is not obvious a priori which financial instrument delivers the best market-based measure of the expected value of the future policy rate.

The general approach described earlier can shed light on this question by ranking the various models according to their forecasting performance where the multiplicity now arises from the multitude of financial instruments (*i.e.* the left-hand-side variable in equation (11)). In fact, Gürkaynak, Sack and Swanson (2007) apply exactly this principle and perform a forecasting horserace to identify the security with the highest predictive power for the federal funds rate. They find that federal funds futures outperform all other instruments for horizons up to six months which they interpret to mean that futures rates provide the best measure for monetary policy expectations; however, they do not account for time-varying risk premia.[18]

Several studies have documented the existence of time-varying risk premia in fed funds futures and their role in distorting the market's assessment of the expected path of monetary policy (see, e.g., Sack, 2004; Piazzesi and Swanson, 2008; Ferrero and Nobili, 2009).[19] For example, Piazzesi and Swanson (2008) show that fed funds (and eurodollar) futures adjusted for a cyclical risk factor produce smaller forecast errors than unadjusted futures, making the risk-adjusted futures the preferred measure of monetary policy expectations. It is likely that time-varying risk premia are an even more relevant component in the other financial market instruments given their different safety and liquidity characteristics which affects the measurement of target rate expectations. Therefore, to derive the most accurate market-based measure of monetary policy expectations a comprehensive analysis is needed that controls for risk premia contained in these different securities and explores additional risk factors.

Inferring market participants' expectations concerning the path of monetary policy poses ad-

[17] Gürkaynak et al. (2007) provide a detailed description of the market characteristics and available maturities for each financial instrument.

[18] They do examine the presence of constant risk premia and find that for many instruments the estimated average risk premia are fairly large; this is consistent with evidence provided in Chernenko et al. (2004).

[19] Hamilton (2009) shows that time-varying risk premia are less of a concern in very short-horizon futures contracts and thus, daily changes in near-dated fed funds futures do primarily reflect the market's expectation of very near-term changes in Fed policy.

ditional challenges when nominal interest rates are close to or at the zero lower bound (ZLB). For example, the fact that in standard Gaussian affine term structure models nothing prevents the short-term rate from turning negative impairs the accuracy of model-implied monetary policy expectations. Bauer and Rudebusch (2016) propose to resort to shadow-rate models that enforce the zero lower bound constraint to recover monetary policy expectations embedded in the yield curve. Their modeling framework produces market-based expectation measures about the path of monetary policy that can be used to determine the timing of policy-rate liftoff and subsequent pace of monetary tightening. They show that including macroeconomic variables as additional risk factors in their shadow-rate model is particularly useful to characterize monetary policy expectations during the ZLB period since yields at the short end lose much of their information content.

Another feature that can influence the size and variation of risk premium estimates are heterogeneous beliefs. For example, Kelly and Pruitt (2013) make the case that exploiting the rich cross-sectional information about individual investors' perceptions of future prices can result in superior estimates of market expectations. Barillas and Nimark (2017, 2019) build a model of the term structure of interest rates in which traders have rational but heterogeneous expectations about future bond prices, in particular the resale value of a bond. They show that the speculative behavior, resulting from the fact that individual traders form expectations based on different subsets of available information, changes the magnitude of historical estimates of time-varying risk premia and expectations about future short rates. In Cao, Crump, Eusepi and Moench (2020), heterogeneity derives from investors' disagreement about the expected path of the policy rate which triggers speculative trading. These differences in beliefs about the level of rates in the long run play an important role in measuring risk premia which impact the market's overall expectation of monetary policy.

4.2 Inflation Expectations

Another interesting case is the derivation of market-based inflation expectations. It might seem that the existence of inflation-linked assets is a precondition for uncovering financial market expectations of inflation. This is not the case. There exist several market-based approaches that have been used to estimate investors' inflation expectations in historical periods that predate the introduction of financial instruments tied to inflation. As long as some market prices incorporate inflation expectations, it is possible to exploit the relationship between observed variables to infer the rates of inflation market participants were anticipating. For example, Hamilton (1985) and Burmeister, Wall and Hamilton (1986) obtain estimates of financial market expectations of inflation by modeling the joint dynamics of nominal interest rates and realized inflation using a state-space framework where expected inflation and real rates are unobserved states. Hamilton (1992) proposes to use the prices of several agricultural commodities traded on futures exchanges to derive expectations about changes in the general price level. The idea is that if there is a stable relationship between commodity prices and consumer prices, then price changes expected by futures markets can be mapped into overall inflation expectations.

While there have been several unsuccessful attempts to establish trading in futures contracts written on the U.S. Consumer Price Index (CPI),[20] other markets for inflation-linked assets have developed during the past two decades, the inflation-indexed bond market and the inflation swap market, both of which convey high-frequency information about investors' views on future inflation.[21] One closely followed indicator is the difference between yields on nominal Treasury securities and Treasury inflation-protected securities (TIPS) for a given maturity, also known as breakeven inflation given that it is the level of inflation that makes investors indifferent between both kinds of securities. The standard practice of treating breakeven inflation rates as pure measures of inflation expectations is problematic, however, since they contain several other components. A salient aspect of breakeven inflation rates is that they involve two securities whose markets differ in important characteristics. It is therefore essential to account for institutional factors that affect nominal and inflation-indexed bonds differently when trying to infer inflation expectation. One major difference is the lower liquidity of TIPS relative to nominal government debt, particularly in periods of financial market stress. Liquidity risks may arise from multiple market frictions, such as limited investor participation, transaction costs, the composition of market participants, funding constraints, as well as net supply imbalances between the two types of securities, for which TIPS investors demand compensation (see, e.g., Pflueger and Viceira, 2011; D'Amico et al., 2018; Andreasen et al., 2020).[22] For example, Pflueger and Viceira (2016) provide empirical evidence for the presence of liquidity premia in inflation-indexed bonds that are economically significant and vary substantially over time. Liquidity issues and other risk factors likely influence the pricing of inflation swaps as well (Campbell et al., 2009; Christensen and Gillan, 2011; Faust and Wright, 2013).

Different empirical strategies have been developed to separate risk factors from the expectation component. Abrahams at al. (2016), D'Amico et al. (2018), Andreasen et al. (2020), among others, use a no-arbitrage pricing framework that jointly models nominal and real yield curves to decompose breakeven inflation rates into expected inflation, inflation risk premia, and liquidity premia. To separately identify these latent components, studies in this literature rely on a variety of pricing factors and modeling choices. For example, D'Amico et al. (2018) model the unobserved liquidity factor as the spread between TIPS yields and the model-implied frictionless real yields and include information on nominal yields, TIPS yields, CPI inflation, and survey-based inflation expectations in the estimation. In contrast, Abrahams et al. (2016) dispense with survey data but use an observable liquidity indicator to adjust TIPS yields for their relative illiquidity. Like D'Amico et al. (2018), Haubrich et al. (2012) use survey forecasts of inflation but replace TIPS yields with inflation swap rates leading them to discard the liquidity factor. Gospodinov and Wei (2018) further explore the promise of financial derivatives as additional sources of market information on

[20] CPI futures contracts were first introduced in June 1985 but trading was suspended in April 1987 due to the lack of investor interest. They were re-introduced in February 2004, restructured in June 2006, and delisted in November 2006. The most recent attempt in 2013 failed even before the contracts were launched.

[21] Campbell, Shiller and Viceira (2009) provide a detailed description and analysis of the evolution of both markets.

[22] Other features that influence the pricing of TIPS include their tax treatment, the seasonality of CPI, the indexation lag, and the embedded deflation protection; see Pflueger and Viceira (2010, 2016), and Christensen and Gillan (2011) for more details. D'Amico et al. (2018) conclude that these factors only play a minor role.

expected inflation by integrating not only inflation swaps but also inflation options and oil futures into an affine term structure framework. Christensen et al. (2010) propose a "yields-only" model with separate level factors for nominal and real yields but common slope and curvature factors. They ignore the liquidity disadvantage of TIPS given that their sample is limited to a period when distortions from liquidity premia were of a lesser concern. Andreasen et al. (2020) use the same basic dynamic factor structure as Christensen et al. (2010) but correct for liquidity risk by exploiting price differentials in the cross section of individual TIPS that arise from the relatively less liquid older securities within the same maturity segment. While the dominant modeling framework for extracting inflation expectations are affine term structure models, there are also a few studies that have used return regressions with various liquidity proxies and risk factors to generate liquidity-adjusted breakeven inflation rates (Gürkaynak et al., 2010; Bauer, 2015; Pflueger and Viceira, 2016) and risk-adjusted inflation swap rates (Casiraghi and Miccoli, 2019).

Again, this results in a situation where there is an abundance of market-based measures of inflation expectations, often with the assertion that a particular measure is better than others with little formal evidence offered in support. While some of these studies show that risk- and liquidity-adjusted breakeven rates improve inflation forecasts compared to unadjusted rates or standard benchmarks like no-change and survey-based forecasts, there is no comparison across different model-implied measures of inflation expectations. To resolve any disagreement between alternative models, it is useful to systematically evaluate the merit of each of these market-based measures by applying the principles laid out in Baumeister and Kilian (2017) to identify the most reliable measure of inflation expectations. While this is straightforward for short- and medium-term inflation expectations, one limitation for long-term inflation expectations is that the available sample might not be long enough yet to evaluate the forecasting performance accurately, as noted by Faust and Wright (2013). To safeguard the selection against spurious forecasting success (until more data accumulate), it seems prudent to pursue a conservative approach that considers the range of implied estimates for market-based measures of long-term inflation expectations across models to obtain lower and upper bounds that can guide monetary policymakers as discussed in Section 5.3 (see also Christensen and Gillan (2011) for a similar approach).

5 Economic Applications of Market-Based Expectation Measures

5.1 Evaluation of Economic Models

Reliable measures of historical price expectations are a precondition for evaluating the empirical content of forward-looking economic models. As pointed out in Section 3.2, it is useful to supplement the statistical model selection with narrative evidence on price developments to increase our confidence in (the accuracy of) market-based expectation measures. I illustrate this by conducting a retrospective analysis of oil price expectations derived from the preferred risk premium model.[23]

[23] The oil price expectation measures can be downloaded from https://sites.google.com/site/cjsbaumeister/research

Panel A of Figure 3 presents the change in the oil price expected over the next year, computed as $E_t(S_{t+12}) - S_t$, where the price expectation is obtained from the HW model for the period 1989.2-2019.6. Between 2005 and mid-2007 the market persistently expected the WTI price to rise by about \$5 within the next year. In the summer of 2008 financial market participants anticipated the oil price to fall by \$22 from its peak of \$133 when a year later it had dropped by half. A strong rebound of more than \$16 to a level of close to \$60 was expected in December 2008 with the realized value one year later a little over \$70. When WTI passed the \$100 mark again in early 2011 the market expected a continuous downward correction over the next 12 months of \$10 on average up until the end of 2014. While the price had been more sluggish to come down relative to what was expected over this period, when its decline accelerated in 2015 the market was optimistic and expected a relatively swift reversal and recovery to about \$60 in 2016/17 which was only reached in early 2018. Thus, the market mostly got the direction of change right, but it did not always anticipate the severity of major events.

Panel B of Figure 3 reports the term structure of futures prices and price expectations of market participants together with the subsequent realized path of the spot oil price at four different points in time, each 18 months apart from the previous one, starting with June 2014.[24] It shows that accounting for the risk premium narrows the gap between what the market expected the price to be and what it turned out to be. However, for some episodes there are still some substantial prediction errors, a question I will turn to in Section 5.2. Overall, the expected price paths are reasonable and in line with historical events lending further credence to the market-based measure.

Next I discuss several settings in which this and other market-based expectation measures can be directly used as inputs for empirical exercises or theoretical models.

Testing Model Hypotheses. Information about price expectations is useful for testing hypotheses. For example, Pavlidis, Paya and Pee (2017) rely on the measure of oil price expectations of Baumeister and Kilian (2017) as an input to test for periodically collapsing speculative bubbles in oil futures markets. The advantage of using market-based expectations in this application is that one does not have to take a stand on how to measure market fundamentals. Their proposed test only requires data on the current and expected spot price. They find no evidence for the existence of bubbles. Studying the evolution of market-based expectations is also helpful to assess the role of rational learning and peso problems in futures markets (e.g., Timmermann, 1993; Leduc, Moran and Vigfusson, 2020). Risk premium estimates across commodity markets can also be informative to test the hypothesis of financialization from 2005 onward. For example, Baumeister et al. (2017) assess the degree of integration among the futures markets for crude oil, motor gasoline, and ethanol based on the correlation of risk premia derived from the HW term structure model estimated for prices of futures contracts in each market. If the increased inflow of index fund investors achieved full market integration, this would be reflected in risk premia being perfectly correlated across markets. Their results do not support the financialization hypothesis.

Model Validation. Expectation measures also serve as an independent source of information to

[24] Baumeister and Kilian (2017) provide evidence for selected dates over the period 2003-2007.

validate modeling assumptions. For example, implicit in the vector autoregressive model used by Wieland (2019) is the assumption that news of oil demand arrives within the period but that oil producers do not respond simultaneously. If oil demand shocks for the next month were known, then producers could adjust production at the time of the change in demand. In that case, the model would attribute forecastable demand shocks to oil supply shocks. Wieland tests if the oil supply shocks are forecastable on the basis of past changes in oil price expectations derived in Baumeister and Kilian (2017). He finds no evidence of predictability which supports the validity of his identification strategy. Reis (2020) exploits the discrepancy between long-run inflation expectations by the public, measured based on household surveys, and by financial market participants, measured based on inflation swaps, to validate models of expectation formation for macroeconomic variables. He shows, for example, that this discrepancy affects inflation dynamics by changing both the effective real interest rate in savings decisions, as well as the nominal interest rate that the central bank chooses to set. He also provides estimates of the underlying expected inflation anchor and illustrates the trade-off that central banks face when deciding how strongly to respond to this discrepancy.

Modeling Storage Demand. The economic theory of storage implies that price spreads are the main determinant of stockpiling (see, e.g., Fama and French, 1987, 1988; Pindyck, 2001). Thus, expected price changes can guide inventory decisions. Baumeister and Kilian (2017) illustrate this for storage in oil markets. When long-term price expectations exceed the current price, traders have an incentive to accumulate inventories as long as this differential is larger than the cost of storage. Instead, when the expected price is below the current price, traders have an incentive to run down their inventories to deal with the implied temporary shortage. Figure 3, panel A indicates that there was a constant incentive to hold inventories throughout most of the 1990s except for the period following the invasion of Kuwait in August 1990 which called for a drawdown of oil stocks. The positive price spread incentivized the built-up of oil inventories from the mid-2000s until the financial crisis of 2008, during which there was a strong incentive to destock. Rising long-term price expectations made storage attractive again in 2009. The period of high oil prices in 2011-13 during which market participants expected oil prices to soften signals the release of inventory holdings. Following the drop in the WTI price in late 2014, however, incentives to hold inventories were restored. Using information about price expectations to gauge incentives for storage is even more useful for commodities for which no (reliable) inventory data exist. For example, Baumeister et al. (2017) extract information about market expectations from ethanol futures prices to test whether storage demand increased in response to changes in biofuel policies.

Modeling Agents' Decisions. Households' purchase decisions of durables and firms' production and investment decisions are forward-looking and thus depend on expected price developments. For example, Kahn (1986), Busse, Knittel and Zettelmeyer (2013), and Allcott and Wozny (2014) analyze how consumers trade off future gasoline costs with the sales price of vehicles with different fuel economy ratings when making car-buying decisions. Kahn (1986) and Allcott and Wozny (2014) show that the choice of the measure for gasoline price expectations matters for determining

whether the market values energy efficiency. To assess the sensitivity of vehicle prices to changes in gasoline price expectations, these studies rely on unadjusted crude oil futures prices and survey data to proxy consumers' forecasts of future gasoline prices. Baumeister et al. (2017) show that risk-adjusted gasoline futures prices outperform both the no-change forecast implied by surveys and 'pure' futures prices which would make them a preferred choice to represent gasoline price expectations in these models. In light of the policy consequences of potential undervaluation of energy costs, in particular the implications for the design of climate policies, it would be interesting to revisit the evidence derived from these models using a superior measure of price expectations. This measure can also be used to estimate how sensitive consumers are to expected operating costs when purchasing other energy-using durable goods. Similarly, market-based commodity price expectations can replace simpler measures when studying resource owners' decisions to extract nonrenewable resources and to invest in the development of new reserves of raw materials (see, e.g., Anderson et al., 2018; Gilje et al., 2020).

5.2 Deriving Shock Measures

Monetary Policy Shocks. Kuttner (2001) forcefully argues that distinguishing between expected and unexpected Federal Reserve policy actions is essential to correctly estimate the impact of policy surprises on the entire yield curve and other asset prices. Key for successfully separating out the unanticipated component of policy decisions is an accurate measure of what the market expected the policy rate to be.

Having identified the single, most accurate measure of near-term target rate expectations, we can apply Rudebusch's (1998) definition of a monetary policy shock as the difference between the realized fed funds rate target and the market-based expectation to obtain the surprise measure. Piazzesi and Swanson (2008) provide an illustration of the difference that risk adjustment of policy expectations makes for the computation of monetary policy shocks. When the measurement of policy surprises is tied to FOMC meetings, as suggested by Kuttner (2001), the timing of policy announcements within the month impacts the derivation of shocks to the immediate policy setting. Therefore, Gürkaynak et al. (2007) advocate constructing policy shocks using expectations with slightly longer horizons to capture changes in the expected near-term policy path which is less affected by shifts in the dates of policy decisions. In the case of using changes in the fed funds futures rate on the day of an FOMC announcement as the policy surprise, risk adjustment might be of lesser importance since risk premia are unlikely to change much at higher frequencies (see, e.g., Kuttner, 2001; Faust et al., 2004; Piazzesi and Swanson, 2008; Hamilton, 2009).

Oil Price Shocks. The oil market is another setting where understanding whether markets expected oil price changes or were surprised by these changes is of immediate interest to policymakers and economists seeking to understand the determinants and economic impact of oil price fluctuations. As in the case of monetary policy, percent deviations of oil price expectations from the realized price of oil can be interpreted as market-based oil price shocks. For example, Baumeister

and Kilian (2016a) examine quarterly expectational shocks to the WTI price for a selected set of historical episodes between 1990 and 2015 to assess how forward-looking financial market participants are relative to policymakers and consumers. Comparing oil price expectations, derived from Brent oil futures using the HW model, to subsequent outcomes from June to December 2014, Baumeister and Kilian (2016b) conclude that investors failed to anticipate the sharp decline in the Brent price over this period, resulting in a large negative oil price shock.

Panel A of Figure 4 plots the time series of oil price surprises computed as the log difference between the realized price of WTI and what market participants expected that price to be last month for the period 1986.1-2020.4. Using surprises defined on monthly intervals makes this market-based shock measure comparable to oil market shocks derived from structural vector autoregressions (see, e.g., Baumeister and Hamilton, 2019). The figure shows that the largest oil price shocks tend to coincide with well-known historical episodes of political and economic turmoil that caught financial market participants by surprise.[25] For example, the market seems to have underestimated the price effects of additional oil supply from Saudi Arabia in early 1986 being surprised by the sharp drop in oil prices within a few months. While traders expected higher oil prices in August 1990, they were surprised on the upside by the magnitude of the price increase as a result of the invasion of Kuwait. In the months after this event, the price reversal happened faster than the market had expected, leading to a series of oil price surprises on the downside. While the market correctly anticipated the price reversal from its all-time high in July 2008, it once again did not foresee the rapid pace of price deterioration in the second half of 2008. In the first few months of 2009, financial market participants were slow to catch on to the price recovery which is reflected in large positive oil price shocks. Over the 2014-16 period, traders were repeatedly surprised by developments in global oil markets both on the upside and the downside, indicating that expectations adapted only gradually. The largest market-based oil price shocks took place in early 2020 when news about the spread of the COVID-19 pandemic made oil prices tumble beyond expectations.

The previous discussion highlights that shocks to market participants' expectations are often related to events that cause shifts in oil demand and oil supply. To gain a better understanding of these shocks, I filter out the 'pure' expectation component by regressing the market-based surprises on fundamental oil supply and demand shocks.[26] Panel B of Figure 4 presents the orthogonalized expectational shocks along with the original oil price surprises to get a sense of the relative importance of 'pure' expectation shocks that are driven by market beliefs that are distinct from new information about fundamentals. Overall, these shocks account for a non-trivial fraction of market-based oil price surprises. It would interesting to examine the macroeconomic dynamics of 'pure' oil price expectation shocks to learn more about their nature. Another possibility would be to include oil price expectation measures directly as an observable in a vector autoregressive (VAR) model

[25] Baumeister and Kilian (2016a) provide a detailed account of the history of oil markets.

[26] Specifically, I use the time series of the four structural oil market shocks of Baumeister and Hamilton (2019), namely oil supply shocks, oil consumption demand shocks, oil inventory demand shocks, and global economic activity shocks. These fundamental shocks together with the market-based oil price surprises and the orthogonalized expectational shocks can be found at https://sites.google.com/site/cjsbaumeister/research.

and isolate the shock component by applying existing identification strategies. For example, Barsky and Sims (2012), D'Amico and King (2017), Levchenko and Pandalai-Nayar (2020), Clements and Galvão (2021), and Lukmanova and Rabitsch (2021) augment standard VARs with survey expectations and study the role of innovations to expectations for business cycle fluctuations and/or the conduct of policy.

So far I have focused on month-to-month oil price surprises. In other contexts, it might be useful to consider shocks to oil price expectations at longer horizons. For example, López-Salido and Loria (2021) use 3-month and 6-month oil price surprises constructed based on the market-based oil price expectation measures of Baumeister and Kilian (2017) to assess the role of oil price shocks for their market-based measure of inflation risks. They show that their options-implied inflation probabilities exhibit a strong correlation with oil price surprises.

5.3 Policy Analysis

Expectations about future price developments play a key role in assessing the effectiveness of public policies, designing appropriate policy instruments, and guiding policy decisions given that they influence the behavior of economic agents.

Regulations and Government Policies. To identify the policy instrument most suitable to achieve a certain policy goal, it is necessary to ensure that consumers or firms respond to incentives intended by the instrument. For example, knowing to what extent consumers factor future fuel costs into their purchase decisions of vehicles is critical for deciding whether fuel-economy regulations (CAFE standards) or price-based policies (gasoline or carbon taxes) are more efficient to curb greenhouse gas emissions (see, e.g., Busse et al., 2013; Allcott and Wozny, 2014). It is also important to know how environmental and other policies affect individual markets. For example, Baumeister et al. (2017) study the implications of the creation of the Renewable Fuel Standard (RFS) for the ethanol market. With the help of market-based price expectations for crude oil and motor gasoline, they design a counterfactual of how future ethanol prices would have evolved in the absence of the policy where the difference between the actual and counterfactual price paths provides an estimate of the causal effects of the RFS. Market-based expectation measures can also be used to inform specific policy interventions. For example, price spreads computed as the difference between the market's expected one-year-ahead oil price and the front-month futures price carry valuable information for the government about when and how much crude oil to release from or add to the U.S. Strategic Petroleum Reserve (see Newell and Prest, 2017).

Monetary Policy. Central banks closely monitor inflation expectations. Among the indicators used to gauge the future path of inflation, market-based measures of expected inflation have gained popularity among policymakers given that they are available at much higher frequencies and for a wider range of time horizons compared to more traditional survey measures. Being able to assess in real time at which horizons relevant changes in the market's inflation outlook take place provides valuable information for conducting monetary policy and evaluating its credibility.

Since monetary policy is forward-looking, near-term inflation expectations are an important input into the decision-making process. Central banks will adjust their policy instrument if expected inflation deviates from a level viewed as optimal to counter those expectations. The higher-frequency nature of market-based inflation expectations is also useful for the purpose of assessing the reaction of expected inflation to conventional and unconventional monetary policy actions using the event study methodology (see, e.g., Abrahams et al., 2016).

Policymakers also pay close attention to long-term inflation expectations to get a sense of the market's confidence in the central bank's ability to achieve its mandate of price stability. Stable levels of long-term inflation expectations are an indication of a central bank's commitment to fight inflation since short-run inflationary pressures due to cyclical factors do not change financial market participants' expectations about the rate of inflation over the longer run (see, e.g., Gürkaynak, Sack and Wright, 2010). Söderlind and Svensson (1997) suggest to use the difference between long-term inflation expectations and a central bank's inflation target as an indicator of investors' perception of the credibility of the monetary policy regime. Knowing whether inflation expectations are well anchored to a central bank's inflation target is also important to obtain a more favorable trade-off between inflation and growth. For example, Gürkaynak, Levin and Swanson (2010) rely on market-based measures of long-run inflation expectations to show that an explicit inflation target helps to more firmly anchor the private sector's expected level of inflation at long horizons which improves economic performance. Other economic benefits of stable and well-anchored inflation expectations include lower variability of long-term nominal interest rates and reduced uncertainty about future inflation (see, e.g., Bauer, 2015).

Market-based real rate expectations can be used to gauge the stance of monetary policy (see, e.g., Christensen and Rudebusch, 2019). This not only allows to assess whether financial market participants perceive current monetary policy as expansionary or contractionary, but also to examine historical episodes; for example, whether the market thought monetary policy was too accommodative during the run-up of house prices in the early-to-mid 2000s.

5.4 Implications for Out-of-Sample Forecasts

My analysis so far has focused on historical market-based price expectations which rely on the information contained in the entire sample. Insofar as risk premia are predictable, it might be interesting to investigate whether risk-adjusting asset prices in a real-time out-of-sample forecasting setting would achieve similar gains in accuracy compared to unadjusted asset prices. It is well known that even if the model is correct, predictive success based on the full sample does not necessarily imply good out-of-sample performance, given the bias-variance trade-off in estimating forecasting models.

I examine this question in the context of forecasting the price of oil. This is useful since, as noted above, oil futures prices are still a popular choice by central banks and international organizations to gauge the future path of oil prices which often feeds into macroeconomic projections and policy

decisions. Pagano and Pisani (2009) were the first to study the promise of risk adjustment of oil futures prices for forecasting purposes. They infer the risk premium based on a U.S. business cycle indicator and show that the risk-adjusted futures price consistently outperforms the unadjusted futures price. Baumeister and Kilian (2017) evaluate the out-of-sample forecast accuracy of the HW term structure model since it produced the most reliable measure of market expectations based on the full-sample selection criterion. They conclude that this new approach to oil price forecasting looks promising and recommend monitoring the evolution of its performance as more data become available given their short evaluation period. I revisit the evidence for an evaluation period that now spans a decade. As shown by Baumeister and Kilian (2017), ignoring the break uncovered by Hamilton and Wu (2014) hurts the forecasting performance of the HW model which is why I focus exclusively on the post-break period. I start the estimation sample in January 2005 and allow for 48 observations in the initial sample which means that the first out-of-sample h-step-ahead forecasts are generated in December 2008.

Column 1 of Table 5 shows the recursive MSPE ratios for the unadjusted futures price for horizons $h = 3, 6, 9, 12$ months ahead evaluated over the period 2009.1 to 2020.7. The shorter evaluation period leads to further improvements in forecast accuracy relative to the no-change forecast across all horizons compared to the evidence presented in Table 4. In particular, it reduces the MSPE by 11% at the 3-month horizon, 16% at the 6-month horizon, 22% at the 9-month horizon, and 26% at the 12-month horizon. Column 2 reports results for futures prices adjusted by the HW risk premium estimated recursively using only data available to the forecaster at the time the forecast was made. While the forecasts still beat the random walk except at horizon 3, the MSPE ratios are on average 16 percentage points higher than those of the unadjusted futures price. The structure of the HW model implies that forecasts are produced at the end of the third week of month t, whereas the convention in out-of-sample forecasting is that we forecast using all of the information available at the end of the current month. To account for this difference in timing, I adjust the level of the HW oil price forecast by the change in the daily futures price of maturity h between the day on which the HW forecast was generated and the last trading day of the month, as suggested by Baumeister and Kilian (2017). The last column of Table 5 shows that this timing adjustment improves the accuracy of the HW forecast by between 2 and 17 percentage points, depending on the horizon, compared to the results in column 2. Yet, also this model produces higher MSPE ratios than the unadjusted futures price at all horizons. This might be surprising given the encouraging performance of the HW model originally reported in Baumeister and Kilian (2017).

Figure 5 reconciles these findings by looking at the evolution of the cumulative mean-squared prediction errors at each horizon for all three models over time. The graphs show that much of the superior forecast accuracy of unadjusted futures prices derives from the post-2014 period. While the timing-adjusted HW model was the best-performing model in the early part of the evaluation sample, the ranking of the models is reversed after the 2014-16 oil price decline. This switch coincides with the end of the evaluation period considered in Baumeister and Kilian (2017) and

thus explains the differences in MSPE ratios which summarize the average forecast accuracy over the two evaluation periods.

This exercise also highlights that the out-of-sample forecasting performance of futures prices has varied considerably over time and might depend on changes in the size of time-varying risk premia.[27] One promising way forward would be to examine the benefits of adding the HW model to the suite of models used for pooling oil price forecasts (see, e.g., Baumeister, 2014; Baumeister and Kilian, 2015). Another approach to combine market-based expectations with time-series models for forecasting purposes is to apply entropic tilting (see, e.g., Robertson, Tallman and Whiteman, 2005; Krüger, Clark and Ravazzolo, 2017). For example, Altavilla, Giacomini and Costantini (2014) use an exponential tilting method to incorporate market expectations extracted from futures contracts on the federal funds rate into model-based forecasts of bond returns. They show that this approach yields sizeable improvements in forecast accuracy and profitable investment strategies for investors in bond markets. More generally, the usefulness of risk-adjusting asset prices for the purposes of out-of-sample forecasting and optimal portfolio choice has not been explored much to date in other contexts and thus offers fruitful avenues for future research.

6 Conclusions

Asset prices are a valuable source of information since they incorporate market participants' expectations about the future. However, they also contain a time-varying risk premium that is unobservable. Thus, to extract information about expectations we need to purge risk premia from asset prices which is easier said than done. In this paper, I illustrated a general approach for recovering market expectations from asset prices and its usefulness in a myriad of economic settings.

[27] Baumeister and Kilian (2017) document that the systematic failure of the oil futures price as a predictor of the spot price between 2003 and mid-2008 can be largely attributed to changes in risk premia.

References

Abrahams, Michael, Tobias Adrian, Richard K. Crump, Emanuel Moench, and Rui Yu (2016). "Decomposing Real and Nominal Yield Curves," *Journal of Monetary Economics* 84: 182-200.

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

Adrian, Tobias, and Hyun Song Shin (2010). "Liquidity and Leverage," *Journal of Financial Intermediation* 19(3): 418-437.

Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant (2018). "Hotelling Under Pressure," *Journal of Political Economy* 126: 984-1026.

Andreasen, Martin M., Jens H.E. Christensen, and Simon Riddell (2020). "The TIPS Liquidity Premium," Federal Reserve Bank of San Francisco Working Paper 2017-11.

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

Allcott, Hunt, and Nathan Wozny (2014). "Gasoline Prices, Fuel Economy, and the Energy Paradox," *Review of Economics and Statistics* 96: 779-795.

Altavilla, Carlo, Raffaella Giacomini, and Riccardo Costantini (2014). "Bond Returns and Market Expectations," *Journal of Financial Econometrics* 12(4): 708-729.

Barillas, Francisco, and Kristoffer P. Nimark (2017). "Speculation and the Term Structure of Interest Rates," *Review of Financial Studies* 30(11): 4003-4037.

Barillas, Francisco, and Kristoffer P. Nimark (2019). "Speculation and the Bond Market: An Empirical No-Arbitrage Framework," *Management Science* 65(9): 4179-4203.

Barsky, Robert B., and Eric R. Sims (2012). "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence," *American Economic Review* 102(4): 1343-1377.

Bauer, Michael D. (2015). "Inflation Expectations and the News," *International Journal of Central Banking* 11(2): 1-40.

Bauer, Michael D., and James D. Hamilton (2018). "Robust Bond Risk Premia," *Review of Financial Studies* 31: 399-448.

Bauer, Michael D., and Glenn D. Rudebusch (2016). "Monetary Policy Expectations at the Zero Lower Bound," *Journal of Money, Credit and Banking* 48(7): 1439-1465.

Baumeister, Christiane (2014). "The Art and Science of Forecasting the Real Price of Oil," *Bank of Canada Review*, Spring: 21-31.

Baumeister, Christiane, Reinhard Ellwanger, and Lutz Kilian (2017). "Did the Renewable Fuel Standard Shift Market Expectations of the Price of Ethanol?" NBER Working Paper 23752.

Baumeister, Christiane, and James D. Hamilton (2019). "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review* 109(5): 1873-1910.

Baumeister, Christiane, and Lutz Kilian (2015). "Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach," *Journal of Business and Economic Statistics* 33(3): 338-351.

Baumeister, Christiane, and Lutz Kilian (2016a). "Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us," *Journal of Economic Perspectives* 30(1): 139-160.

Baumeister, Christiane, and Lutz Kilian (2016b). "Understanding the Decline in the Price of Oil Since June 2014," *Journal of the Association of Environmental and Resource Economists* 3(1): 131-158.

Baumeister, Christiane, and Lutz Kilian (2017). "A General Approach to Recovering Market Expectations from Futures Prices With an Application to Crude Oil," CEPR Discussion Paper 10162.

Baumeister, Christiane, Dimitris Korobilis, and Thomas K. Lee (2020). "Energy Markets and Global Economic Conditions," *Review of Economics and Statistics*, forthcoming.

Bekaert, Geert, and Robert J. Hodrick (1993). "On Biases in the Measurement of Foreign Exchange Risk Premiums," *Journal of International Money and Finance* 12: 115-138.

Bernanke, Ben S. (2008). "Outstanding Issues in the Analysis of Inflation," Speech at the Federal Reserve Bank of Boston's 53rd Annual Economic Conference, Chatham, Massachusetts on June 9.

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

Bessembinder, Hendrik, and Kalok Chan (1992). "Time-Varying Risk Premia and Forecastable Returns in Futures Markets," *Journal of Financial Economics* 32: 169-193.

Bessembinder, Hendrik, and Paul J. Seguin (1993). "Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets," *Journal of Financial and Quantitative Analysis* 28(1): 21-39.

Burmeister, Edwin, Kent D. Wall, and James D. Hamilton (1986). "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering," *Journal of Business and Economic Statistics* 4(2): 147-160.

Busse, Meghan, Christopher R. Knittel, and Florian Zettelmeyer (2013). "Are Consumers Myopic? Evidence from New and Used Car Purchases," *American Economic Review* 103: 220-256.

Campbell, John Y., and John H. Cochrane (1999). "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107(2): 205-251.

Campbell, John Y., and Robert J. Shiller (1991). "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies* 58: 495-514.

Campbell, John Y., Robert J. Shiller, and Luis M. Viceira (2009). "Understanding Inflation-Indexed Bond Market," *Brookings Papers on Economic Activity*, Spring: 79-120.

Cao, Shuo, Richard K. Crump, Stefano Eusepi, and Emanuel Moench (2020). "Fundamental Disagreement about Monetary Policy and the Term Structure of Interest Rates," CEPR Discussion Paper 15122.

Casiraghi, Marco, and Marcello Miccoli (2019). "Inflation Risk Premia and Risk-Adjusted Expectations of Inflation," *Economics Letters* 175: 36-39.

Chernenko, Sergey V., Krista B. Schwarz, and Jonathan H. Wright (2004). "The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk," International Finance Discussion Paper 808, Board of Governors of the Federal Reserve System.

Chinn, Menzie D., and Olivier Coibion (2014). "The Predictive Content of Commodity Futures," *Journal of Futures Markets* 34(7): 607–636.

Christensen, Jens H.E., and James Gillan (2011). "TIPS Liquidity, Breakeven Inflation, and Inflation Expectations," *FRBSF Economic Letter* 2011-19.

Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch (2010). "Inflation Expectations and Risk Premiums in an Arbitrage-Free Model of Nominal and Real Bond Yields," *Journal of Money, Credit and Banking* 42(6): 143-178.

Christensen, Jens H.E., and Glenn D. Rudebusch (2019). "A New Normal for Interest Rates? Evidence from Inflation-Indexed Debt," *Review of Economics and Statistics* 101(5): 933-949.

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

Clements, Michael P., and Ana Beatriz Galvão (2021). "Measuring the Effects of Expectations Shocks," *Journal of Economic Dynamics and Control* 124: 104075.

Cochrane, John H. (2011). "Presidential Address: Discount Rates," *Journal of Finance* 66(4): 1047–1108.

Cochrane, John H., and Monika Piazzesi (2005). "Bond Risk Premia," *American Economic Review* 95(1): 138-160.

Dai, Qiang, and Kenneth J. Singleton (2002). "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics* 63(3): 415-441.

D'Amico, Stefania, Don H. Kim, and Min Wei (2018). "Tips from TIPS: The Informational Content of Treasury Inflation-Protected Security Prices," *Journal of Financial and Quantitative Analysis* 53(1): 395-436.

D'Amico, Stefania, and Thomas King (2017). "What Does Anticipated Monetary Policy Do?" Federal Reserve Bank of Chicago, mimeo.

Deaves, Richard, and Itzhak Krinsky (1992). "Risk Premiums and Efficiency in the Market for Crude Oil Futures," *Energy Journal* 13(2): 93-117.

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

Dincerler, Cantekin, Zeigham Khokher, and Timothy Simin (2020). "An Empirical Analysis of Commodity Convenience Yields," *Quarterly Journal of Finance* 10(2): 1-27.

Dusak, Katherine (1973). "Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums," *Journal of Political Economy* 81(6): 1387–1406.

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

Fama, Eugene F. (1984). "Forward and Spot Exchange Rates," *Journal of Monetary Economics* 14: 319-338.

Fama, Eugene F. (2014). "Two Pillars of Asset Pricing," *American Economic Review* 104(6): 1467-1485.

Fama, Eugene F., and Robert R. Bliss (1987). "The Information in Long-Maturity Forward Rates," *American Economic Review* 77(4): 680-692.

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

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

Faust, Jon, Eric T. Swanson, and Jonathan H. Wright (2004). "Identifying VARS Based on High Frequency Futures Data," *Journal of Monetary Economics* 51: 1107-1131.

Faust, Jon, and Jonathan H. Wright (2013). "Forecasting Inflation," in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2A, Amsterdam: North-Holland, 3-56.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020). "Taming the Factor Zoo: A Test of New Factors," *Journal of Finance* 75(3): 1327-1370.

Ferrero, Giuseppe, and Andrea Nobili (2009). "Futures Contract Rates as Monetary Policy Forecasts," *International Journal of Central Banking* 5(2): 109-146.

Ferson, Wayne E., and Campbell R. Harvey (1991). "The Variation of Economic Risk Premiums," *Journal of Political Economy* 99(2): 385-415.

Froot, Kenneth A. (1989). "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Finance* 44(2): 283-305.

Froot, Kenneth A. and Jeffrey A. Frankel (1989). "Forward Discount Bias: Is it an Exchange Risk Premium?" *Quarterly Journal of Economics* 104(1): 139–161.

Gilje, Erik P., Elena Loutskina, and Daniel Murphy (2020). "Drilling and Debt," *Journal of Finance* 75(3): 1287-1325.

Gospodinov, Nikolay, and Bin Wei (2018). "Forecasts of Inflation and Interest Rates in No-Arbitrage Affine Models," Federal Reserve Bank of Atlanta, mimeo.

Gürkaynak, Refet S., Andrew Levin, and Eric T. Swanson (2010). "Does Inflation Targeting Anchor Long-Run Inflation Expectations? Evidence from Long-Term Bond Yields in the U.S., U.K., and Sweden," *Journal of the European Economic Association* 8(6): 1208-42.

Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson (2005). "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review* 95: 425-436.

Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson (2007). "Market-Based Measures of Monetary Policy Expectations," *Journal of Business and Economic Statistics* 25(2): 201-212.

Gürkaynak, Refet S., Brian P. Sack, and Jonathan H. Wright (2007). "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics* 54(8): 2291-2304.

Gürkaynak, Refet S., Brian P. Sack, and Jonathan H. Wright (2010). "The TIPS Yield Curve and Inflation Compensation," *American Economic Journal: Macroeconomics* 2(1): 70-92.

Gürkaynak, Refet S., and Jonathan H. Wright (2012). "Macroeconomics and the Term Structure," *Journal of Economic Literature* 50(2): 331-67.

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

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

Hamilton, James D. (1985). "Uncovering Financial Market Expectations of Inflation," *Journal of Political Economy* 93(6): 1224-1241.

Hamilton, James D. (1992). "Was the Deflation During the Great Depression Anticipated? Evidence from the Commodity Futures Market," *American Economic Review* 82(1): 157-178.

Hamilton, James D. (2009). "Daily Changes in Fed Funds Futures Prices," *Journal of Money, Credit and Banking* 41(4): 567-582.

Hamilton, James D., and Tatsuyoshi Okimoto (2011). "Sources of Variation in Holding Returns for Fed Funds Futures Contracts," *Journal of Futures Markets* 31(3): 205-229.

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

Hamilton, James D., and Cynthia J. Wu (2015). "Effects of Index-Fund Investing on Commodity Futures Prices," *International Economic Review* 56: 187-205.

Hansen, Lars P., and Robert J. Hodrick (1988). "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88(5): 829-853.

Harvey, Campbell R., and Yan Liu (2019). "A Census of the Factor Zoo," working paper, Duke University.

Haubrich, Joseph G., George Pennacchi, and Peter H. Ritchken (2012). "Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps," *Review of Financial Studies* 25(5): 1588-1629.

Kahn, James A. (1986). "Gasoline Prices and the Used Automobile Market: A Rational Expectations Asset Price Approach," *Quarterly Journal of Economics* 101: 323-339.

Kelly, Bryan, and Seth Pruitt (2013). "Market Expectations in the Cross-Section of Present Value," *Journal of Finance* 68(5): 1721-1756.

Korajczyk, Robert A. (1985). "The Pricing of Forward Contracts for Foreign Exchange," *Journal of Political Economy* 93(2): 346-368.

Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo (2017). "Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts," *Journal of Business and Economic Statistics* 35(3): 470-485.

Kuttner, Kenneth N. (2001). "Monetary Policy Surprises and Interest Rates: Evidence from the Federal Funds Futures Market," *Journal of Monetary Economics* 47: 523-544.

Leduc, Sylvain, Kevin Moran, and Robert J. Vigfusson (2020), "Learning in the Oil Futures Markets: Evidence and Macroeconomic Implications," *Review of Economics and Statistics*, forthcoming.

Levchenko, Andrei A., and Nitya Pandalai-Nayar (2020). "TFP, News, and "Sentiments:" The International Transmission of Business Cycles," *Journal of the European Economic Association* 18(1): 302-341.

López-Salido, David, and Francesca Loria (2021). "Inflation at Risk," Federal Reserve Board, mimeo.

Ludvigson, Sydney C., and Serena Ng (2009). "Macro Factors in Bond Risk Premia," *Review of Financial Studies* 22(12): 5027-5067.

Lukmanova Elizaveta, and Katrin Rabitsch (2021). "New Evidence on Monetary Transmission: Interest Rate versus Inflation Target Shocks," working paper, KU Leuven.

Newell, Richard G., and Brian C. Prest (2017). "Informing SPR Policy through Futures and Inventory Dynamics," NBER Working Paper 23974.

Newey, Whitney K., and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55(3): 703-708.

Pagano, Patrizio, and Massimiliano Pisani (2009). "Risk-Adjusted Forecasts of Oil Prices," *B.E. Journal of Macroeconomics* 9(1): 1-25.

Pavlidis, Efthymios, Ivan Paya, and David A. Peel (2018). "Using Market Expectations to Test for Speculative Bubbles in the Crude Oil Market," *Journal of Money, Credit, and Banking* 50(5): 833-856.

Pflueger, Carolin E., and Luis M. Viceira (2011). "Inflation-Indexed Bonds and the Expectations Hypothesis," *Annual Review of Financial Economics* 3(1): 139-158.

Pflueger, Carolin E., and Luis M. Viceira (2016). "Return Predictability in the Treasury Market: Real Rates, Inflation, and Liquidity," in: P. Veronesi (ed.), *Handbook of Fixed-Income Securities*, Wiley, NJ, chapter 10.

Piazzesi, Monika (2010). "Affine Term Structure Models," in: Y. Aït-Sahalia and L. Hansen (eds.), *Handbook of Financial Econometrics: Tools and Techniques*, 1, North-Holland, 691-766.

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

Pindyck, Robert S. (2001). "The Dynamics of Commodity Spot and Futures Markets: A Primer," *Energy Journal* 22(3): 1-29.

Reis, Ricardo (2020). "The People versus the Markets: A Parsimonious Model of Inflation Expectations," working paper, LSE.

Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2005). "Forecasting Using Relative Entropy," *Journal of Money, Credit and Banking* 37(3): 383-401.

Rudebusch, Glenn D. (1998). "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review* 39: 907-931.

Rudebusch, Glenn D., and Tao Wu (2008). "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy," *Economic Journal* 118(530): 906-926.

Sack, Brian P. (2004). "Extracting the Expected Path of Monetary Policy from Futures Rates," *Journal of Futures Markets* 24(8): 733-754.

Söderlind, Paul, and Lars Svensson (1997). "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics* 40: 383-429.

Stambaugh, Robert F. (1988). "The Information in Forward Rates: Implications for Models of the Term Structure," *Journal of Financial Economics* 21: 41-70.

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

Timmermann, Allan, and Clive W.J. Granger (2004). "Efficient Market Hypothesis and Forecasting," *International Journal of Forecasting* 20: 15-27.

Wieland, Johannes F. (2019). "Are Negative Supply Shocks Expansionary at the Zero Lower Bound?" *Journal of Political Economy* 127(3): 973-1007.

Table 1. Forecast Efficiency Regressions for Energy Products

Monthly horizon	$\hat{\alpha}$	$\hat{\beta}$	$H_0: \beta = 1$	$H_0: \alpha = 0, \beta = 0$	$H_0: \alpha = 0, \beta = 1$	Adjusted R^2	N
WTI crude oil							
3	0.017*	0.838***	0.570	0.010	0.087	0.023	393
6	0.048**	0.985***	0.965	0.005	0.047	0.037	390
9	0.073**	0.750**	0.396	0.004	0.026	0.051	387
12	0.094**	0.808**	0.560	0.005	0.068	0.067	350
Brent crude oil							
3	0.020*	0.743**	0.458	0.054	0.037	0.012	364
6	0.051**	0.876**	0.768	0.047	0.038	0.039	342
9	0.073**	0.636*	0.333	0.067	0.030	0.026	322
12	0.117**	0.800*	0.625	0.034	0.049	0.049	285
Gasoline							
3	0.034***	1.184***	0.193	0.000	0.001	0.221	387
6	0.069***	1.175***	0.185	0.000	0.001	0.279	380
9	0.117***	1.339***	0.080	0.000	0.002	0.253	273
12	0.140***	1.343***	0.198	0.000	0.014	0.178	262
Heating oil							
3	0.020**	0.800***	0.198	0.000	0.010	0.061	387
6	0.046**	0.724***	0.159	0.001	0.004	0.062	384
9	0.071**	0.715***	0.247	0.004	0.015	-0.011	381
12	0.092**	0.977***	0.943	0.002	0.136	0.094	341
Natural gas							
3	-0.006	0.984***	0.877	0.000	0.864	0.207	342
6	0.006	0.758***	0.020	0.000	0.066	0.169	339
9	0.010	0.872***	0.335	0.000	0.626	0.179	336
12	0.033	0.738***	0.166	0.000	0.374	0.103	331

NOTES: One, two, and three asterisks indicate whether α and β are significantly different from zero at the 10%, 5%, and 1% levels, respectively. All *t*-tests and Wald tests have been computed based on Newey-West heteroscedasticity- and autocorrelation-consistent (HAC) standard errors with lag truncation parameter set equal to the number of overlapping observations. For the three null hypotheses we report asymptotic *p*-values. The end date is 2018.12 and is common across all specifications. The start date is specific to the commodity and the horizon (see Table 1A). *N* is the number of monthly observations.

Table 2. Forecast Efficiency Regressions for Base Metals

Monthly horizon	$\hat{\alpha}$	$\hat{\beta}$	$H_0: \beta = 1$	$H_0: \alpha = 0, \beta = 0$	$H_0: \alpha = 0, \beta = 1$	Adjusted R^2	N
Copper							
3	0.015	-0.377	0.025	0.322	0.050	-0.029	255
6	0.040	-0.601	0.192	0.266	0.191	-0.002	252
9	0.065*	-0.799	0.205	0.228	0.156	-0.051	249
12	0.085*	-1.202	0.122	0.206	0.099	-0.037	246
Lead							
3	0.016*	-0.602	0.002	0.173	0.005	-0.004	255
6	0.045*	-0.780	0.015	0.202	0.033	0.002	252
9	0.080*	-1.078	0.030	0.171	0.059	-0.066	249
12	0.118**	-0.951	0.027	0.141	0.049	0.010	243
Nickel							
3	0.014	0.059	0.292	0.508	0.317	-0.017	255
6	0.046	0.233	0.547	0.404	0.293	-0.011	252
9	0.091	0.540	0.708	0.298	0.236	-0.010	248
12	0.141*	0.889	0.912	0.215	0.222	-0.088	246
Tin							
3	0.015*	0.270	0.345	0.199	0.172	-0.007	255
6	0.038*	-0.649	0.138	0.169	0.049	-0.006	252
9	0.066*	-0.854	0.149	0.120	0.033	-0.004	248
12	0.094	-0.911	0.224	0.129	0.045	-0.002	243
Zinc							
3	-0.002	1.750**	0.370	0.056	0.620	0.020	255
6	0.013	1.658	0.537	0.083	0.537	0.010	252
9	0.043	1.351	0.738	0.099	0.471	-0.012	249
12	0.077	1.284	0.823	0.053	0.332	-0.029	246

NOTES: See Table 1.

Table 3. Monthly Predictor Variables for Payoff on WTI Oil Futures Contracts

Article	Model	Predictors
Bessembinder (1992)	B1	CRSP value-weighted equity index returns
	B2	CRSP value-weighted equity index returns
		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)
Bessembinder and Seguin (1993)	BS	Ratio of trading volume of oil futures contracts to open interest by horizon
Fama and French (1987)	FF1	Constant risk premium
	FF2	Horizon-specific basis (time-varying risk premium)
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	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)
		Horizon-specific basis
	HY2	HY1 + growth rate of dollar open interest for oil futures
	HY3	HY1 + CFNAI
	HY4	HY3 + growth rate of dollar open interest for oil futures
	HY5	HY1 + futures market imbalance
	HY6	HY5 + growth rate of dollar open interest for oil futures
	HY7	HY5 + CFNAI
	HY8	HY7 + growth rate of dollar open interest for oil futures
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
Pagano and Pisani (2009) Extensions	PPE1	GECON from Baumeister, Korobilis, and Lee (2020)
	PPE2	PP2 + GECON
	PPE3	PP3 + GECON

NOTES: The sample period for all variables is 1997.8-2018.12.

Table 4. Predictive Accuracy of Risk-Adjusted Futures Prices for WTI Crude Oil MSPE Ratios for Evaluation Period: 1997.8-2018.12

	Monthly horizon h			
Models	3	6	9	12
F_t^h	**0.976***	**0.965****	**0.923****	**0.859****
Basis Regressions				
FF1	1.013	1.037	1.027	**0.985***
FF2	1.015	1.036	1.029	**0.987***
Payoff Regressions				
B1	**0.984***	1.022	1.017	**0.975***
B2	**0.899***	**0.930****	**0.931****	**0.865****
BC	**0.994**	1.020	1.005	**0.959***
BS	1.003	1.004	1.055	1.016
DNV1	**0.925****	**0.978**	**0.938***	**0.853****
DNV2	**0.925****	**0.969**	**0.939***	**0.850****
GHR1	**0.957****	**0.989***	1.031	**0.994***
GHR2	1.011	1.037	1.015	**0.980***
HY1	**0.977****	**0.992**	**0.989**	**0.938***
HY2	**0.975***	**0.995**	**0.993**	**0.947***
HY3	**0.909****	**0.955****	**0.963****	**0.915****
HY4	**0.912****	**0.954****	**0.963****	**0.926****
HY5	**0.970***	**0.954**	**0.906***	**0.848****
HY6	**0.972***	**0.957**	**0.908***	**0.861****
HY7	**0.887****	**0.894****	**0.849****	**0.794****
HY8	**0.892****	**0.888****	**0.838****	**0.801****
PP1	1.003	1.031	1.032	**0.997***
PP2	**0.979***	**0.989***	**0.981***	**0.960***
PP3	1.012	1.013	**0.949****	**0.865****
PPE1	**0.953***	**0.995***	**0.986***	**0.942****
PPE2	**0.926****	**0.945****	**0.943****	**0.930****
PPE3	**0.954***	**0.973****	**0.907****	**0.825****
Term Structure Model				
HW	**0.896***	**0.829****	**0.762****	**0.697****

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 MSPE reductions are denoted ** at the 5% level and * at the 10% level based on the tests of Diebold and Mariano (1995) and Clark and West (2007), as appropriate. The underlying risk-premium estimates are based on the full sample. HW refers to Hamilton and Wu (2014) and the other labels correspond to the model specifications listed in Table 3. Red indicates the model with the lowest MSPE ratio at each horizon.

Table 5. Recursive MSPE Ratios Relative to No-Change Forecast of the WTI Oil Price
Evaluation Period: 2009.1-2020.7

Monthly horizon h	F_t^h	HW	HW + daily price change
3	**0.890****	1.066	**0.901****
6	**0.840****	**0.972***	**0.935****
9	**0.781****	**0.945****	**0.909****
12	**0.739****	**0.916****	**0.894****

NOTES: Boldface indicates improvements on the monthly no-change forecast. Statistically significant MSPE reductions are denoted ** at the 5% level and * at the 10% level based on the tests of Diebold and Mariano (1995) and Clark and West (2007), as appropriate. The HW term structure model estimates are based on data starting in 2005.1 (after the break date identified by Hamilton and Wu (2014)). HW + daily price change adds to the HW forecast the change in the daily oil futures price of maturity h between the day on which the forecast is generated by the HW model (around the 20th day of the current month) and the last trading day of that month.

Figure 1. Slope Coefficient β Adjusted for Expectational Errors γ as Function of the Relative Standard Deviation σ for Selected Correlation Coefficients ρ

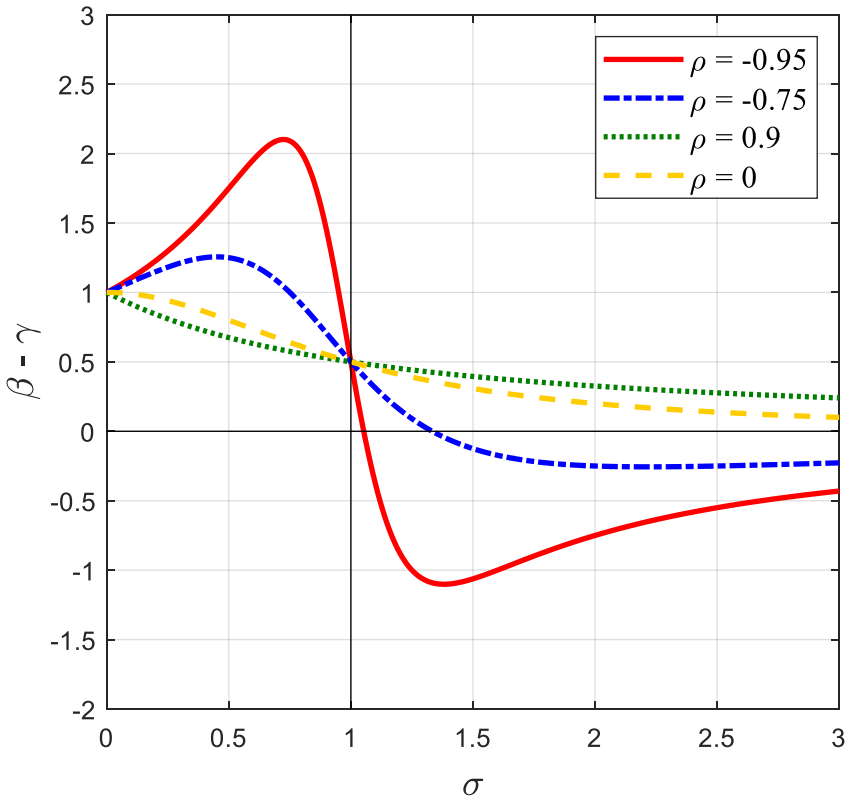
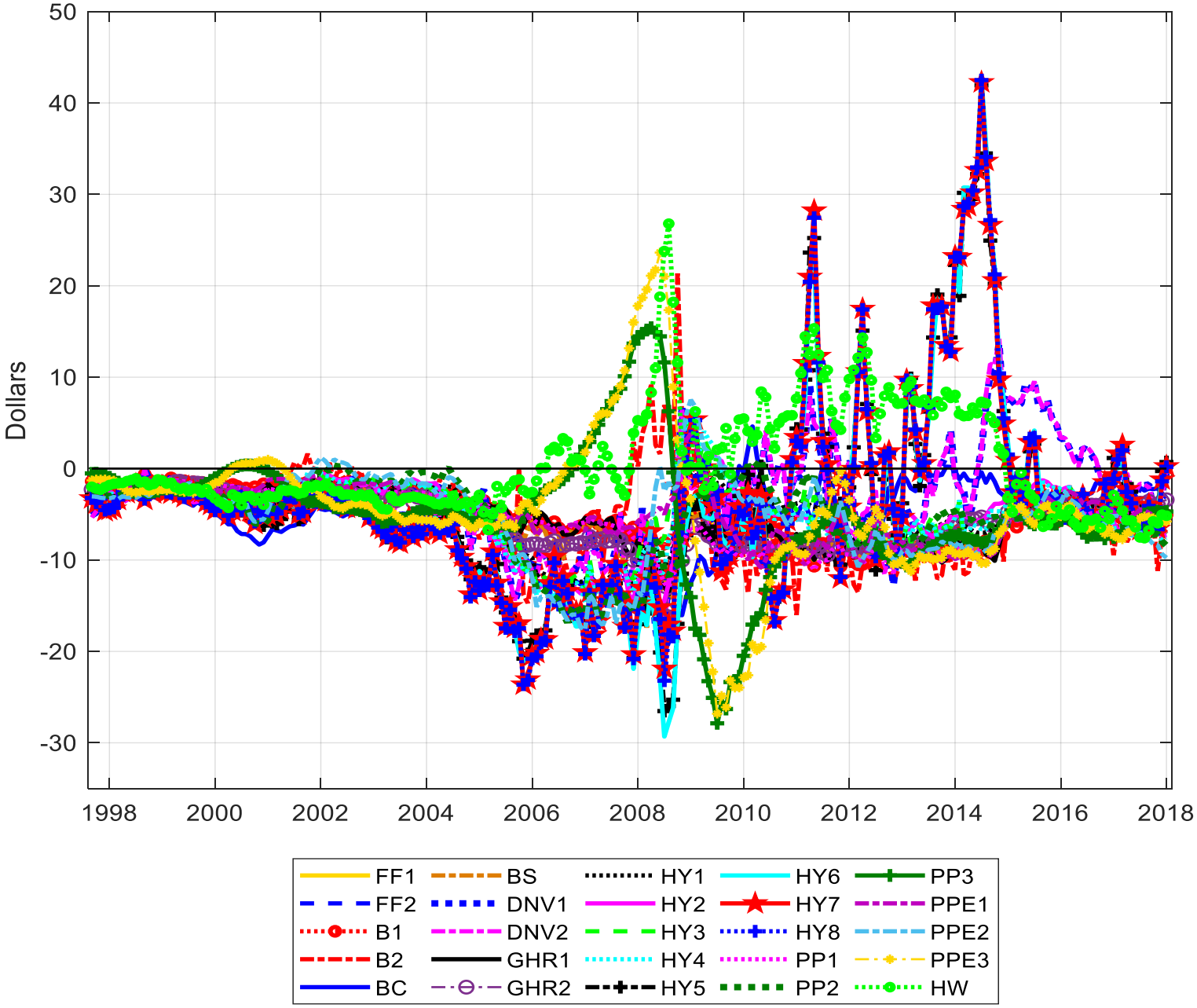


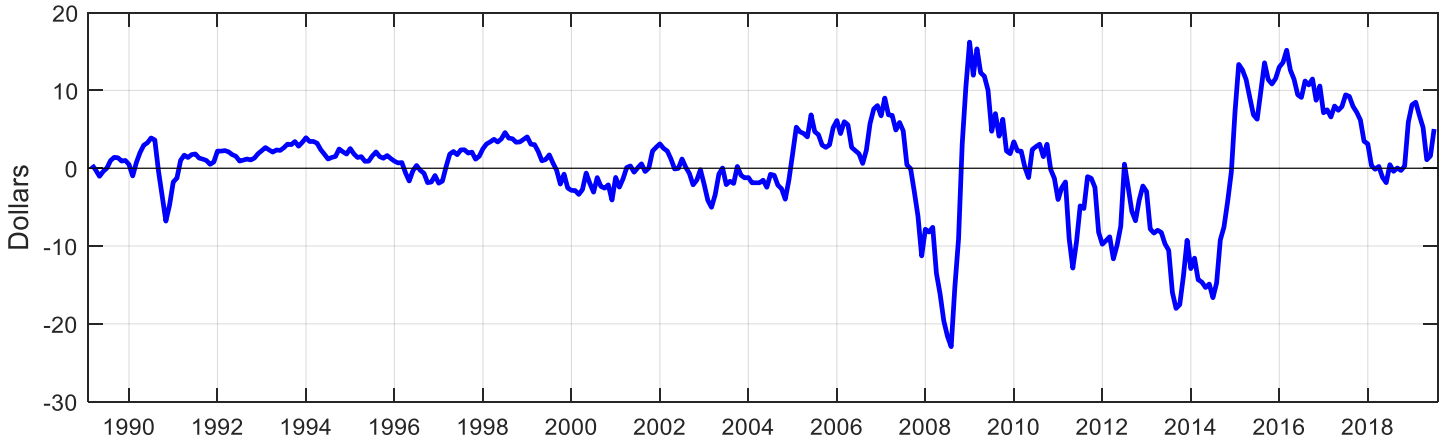
Figure 2. Alternative Monthly Estimates of the Time-Varying Risk Premium in the WTI Oil Futures Market at the 12-Month Horizon 1997.8-2017.12



NOTES: The graph shows 25 alternative estimates of the time-varying risk premium based on the models summarized in Table 3. Qualitatively similar results are obtained for other horizons.

Figure 3. Historical Oil Price Expectations

Panel A: Expected change in WTI crude oil price over the next year, 1989.2-2019.6



Panel B: Trajectories of the WTI futures price F_t^h, the risk-adjusted futures price $F_t^h - RP_t^{HW}$ and the realized spot price S_{t+h} for selected episodes

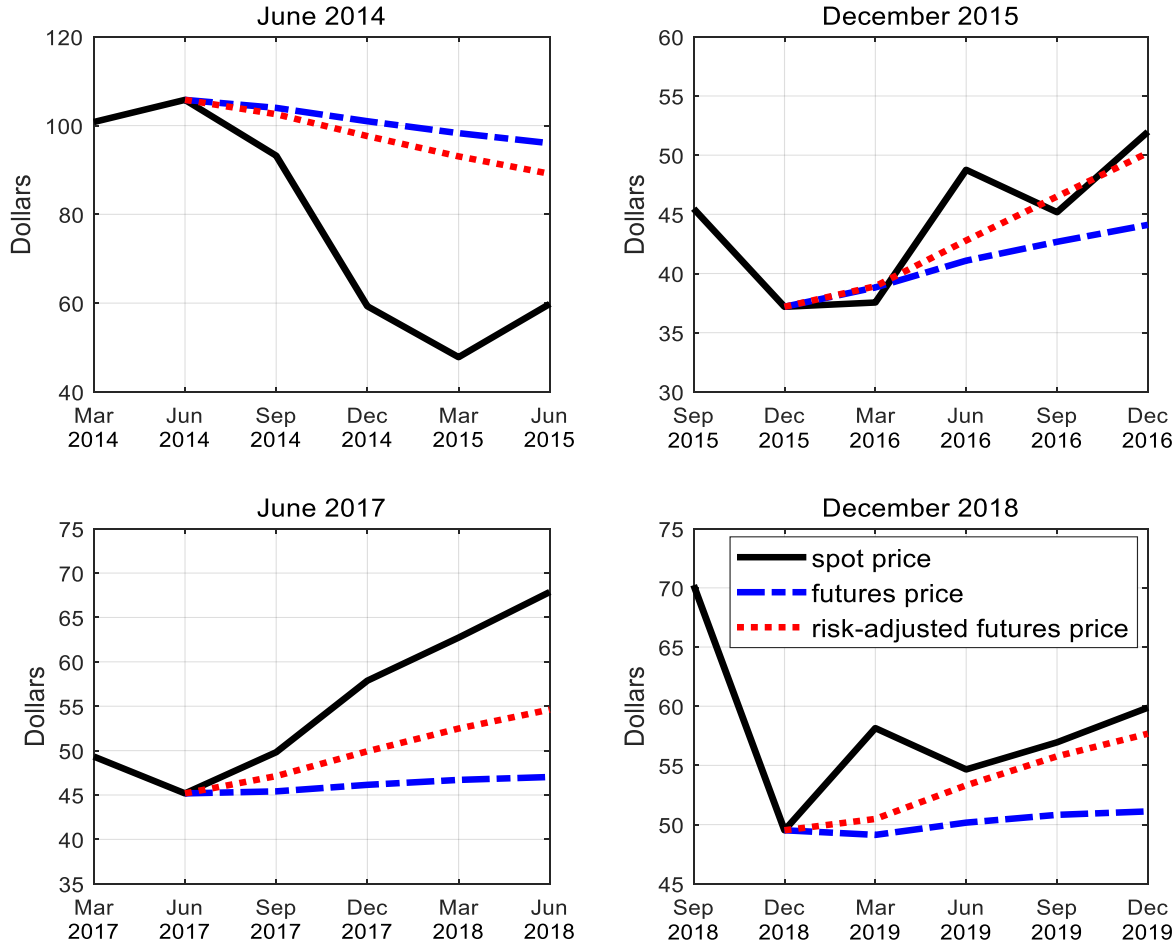
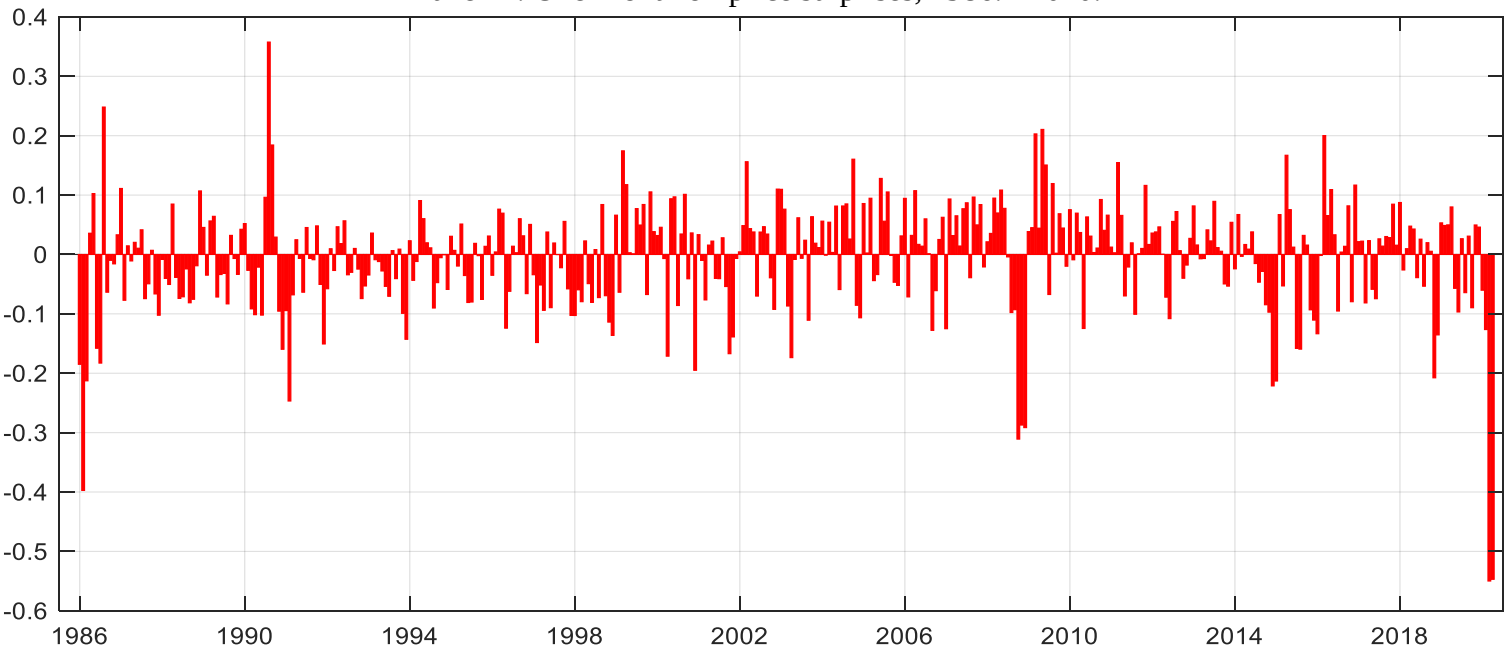


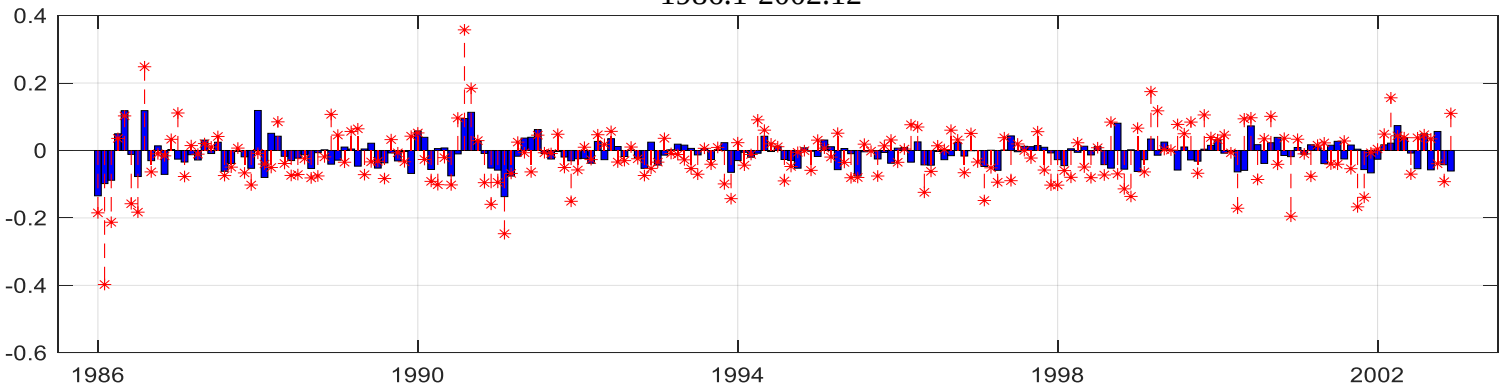
Figure 4. Market-Based Oil Price Shocks

Panel A: One-month oil price surprises, 1986.1-2020.4

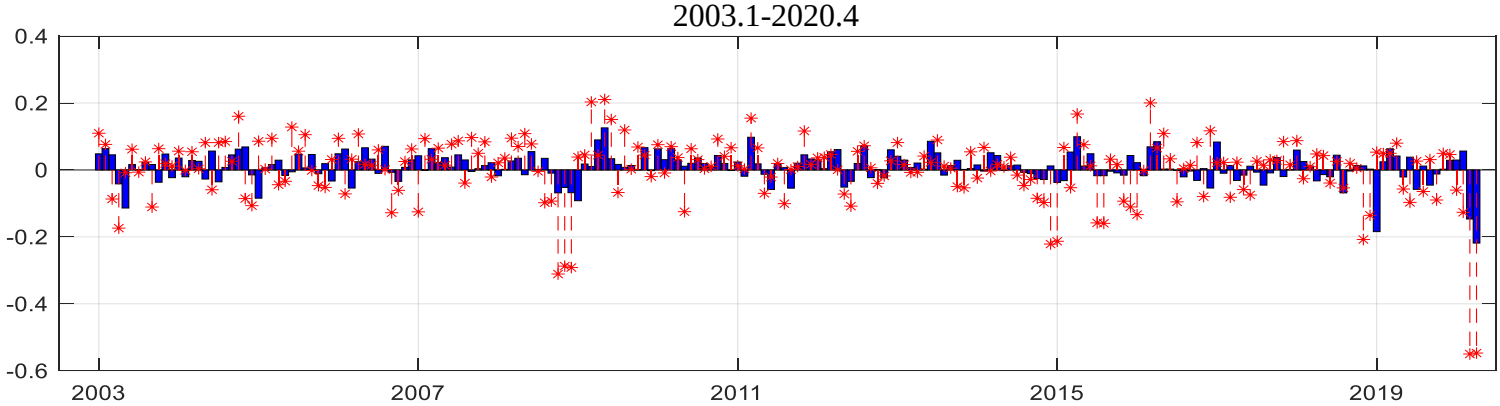


Panel B: 'Pure' oil price expectation shocks

1986.1-2002.12



2003.1-2020.4



NOTES: The red bars in panel A and the red dashed lines with a star in panel B are the log difference between the realized WTI price and the oil price that was expected last month. The blue bars in panel B are 'pure' expectational shocks orthogonal to the four fundamental oil market shocks in Baumeister and Hamilton (2019).

Figure 5. Cumulative Out-of-Sample Mean-Squared Prediction Errors for WTI Forecasts
Evaluation Period: 2009.1-2020.7

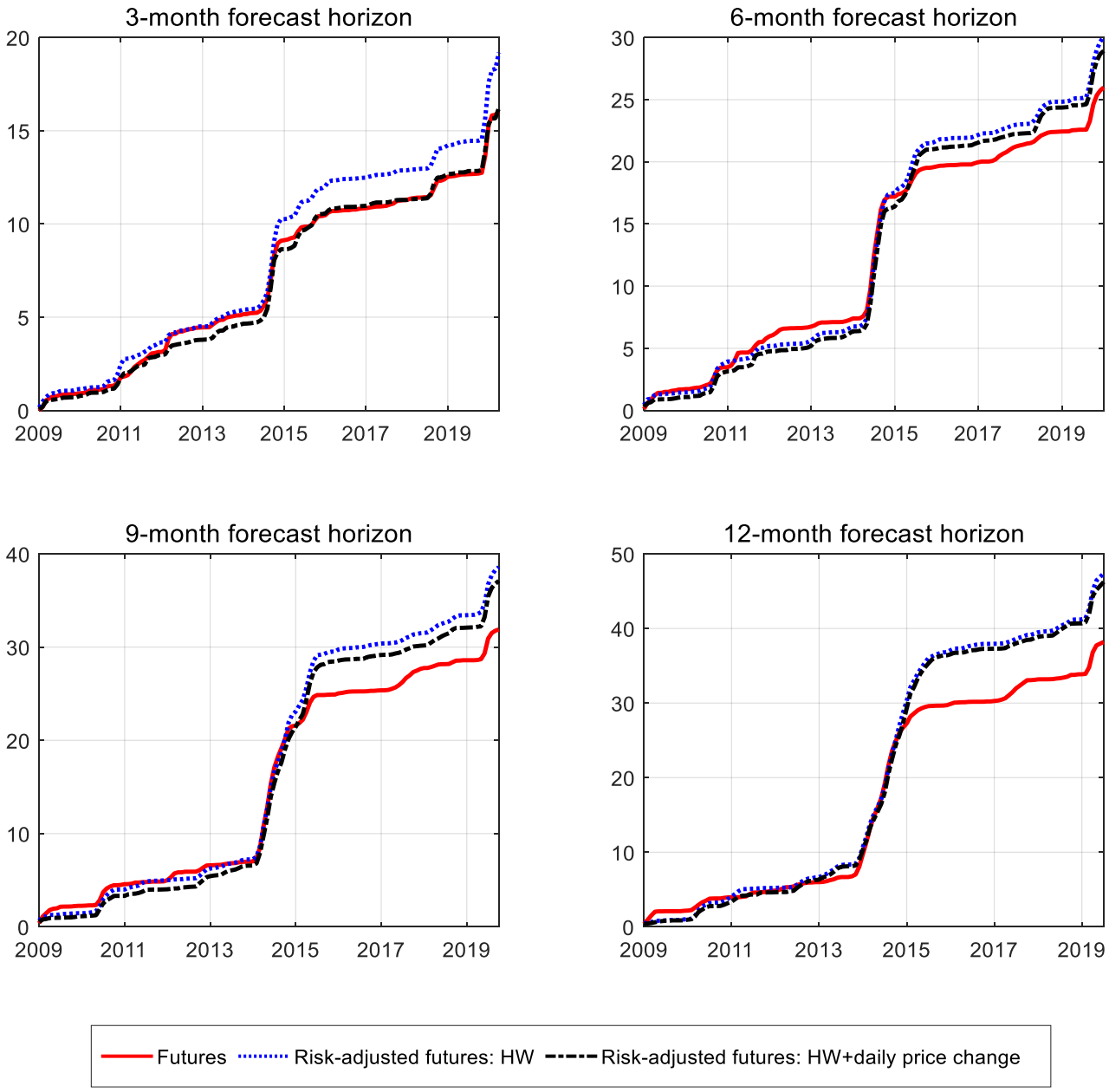


Table 1A. Data Sources and Start Date of Sample Period

	Commodity exchange	Bloomberg futures ticker	First observation for futures contracts with maturity			
			3 months	6 months	9 months	12 months
Energy Products						
WTI crude oil	NYMEX	CL	1986.1	1986.1	1986.1	1988.11
Brent crude oil	NYMEX	CO	1988.6	1990.1	1991.6	1994.4
Gasoline	NYMEX	HU/RB	1986.7	1986.11	1995.7	1996.3
Heating oil	NYMEX	HO	1986.7	1986.7	1986.7	1989.8
Natural gas	NYMEX	NG	1990.4	1990.4	1990.4	1990.6
Base Metals						
Copper	LME	HG	1997.7	1997.7	1997.7	1997.7
Lead	LME	LL	1997.7	1997.7	1997.7	1997.10
Nickel	LME	LN	1997.7	1997.7	1997.8	1997.7
Tin	LME	LT	1997.7	1997.7	1997.8	1997.10
Zinc	LME	LX	1997.7	1997.7	1997.7	1997.7

NOTES: NYMEX refers to the New York Mercantile Exchange and LME refers to the London Metals Exchange. The futures contract for conventional regular unleaded gasoline (HU) for delivery in New York Harbor ceased trading in January 2007. In October 2005 it was replaced by a gasoline futures contract for reformulated blendstock for oxygenate blending (RB) for delivery in New York Harbor. We use the HU prices until December 2005 and the RB prices from January 2006 onwards. The last observation for all commodity futures prices is 2018.12.