The Information Content of Futures Prices: Theory and Case Studies

We examine whether oil futures prices – a highly-liquid, globally utilized commodity – contain information about the future path of retail gasoline prices and atmospheric carbon dioxide. Our work builds on the ideas of Roll (1984), who showed that orange juice futures could help improve weather forecasts. Using modern statistical methods, we found that oil futures have significant predictive power for retail gasoline prices, but less so for the carbon dioxide content in the atmosphere.

Can futures prices predict the future? We can expect futures prices to be informative through two main channels. First, crude oil prices represent more than half the cost of retail gasoline (1). Given the widespread use of crude oil futures in retail gasoline, the retail gasoline market is expected to react to changes in futures prices. If those reactions occur with some lag, then the changes in crude oil futures prices can be predictive of retail gasoline prices.

Second, embedded in futures prices is an implied cost of carry – the market's assessment of the costs and benefits associated with possessing and managing a commodity over a specific time period. In theory, futures prices are the result of no-arbitrage constraints and depend solely on the *cost of carry*. In the case of futures on commodities, the cost of carry is broken down into a *convenience yield*, a *cost of storage*, and the risk-less rate. Explicitly, for a commodity which, between times t and T, has a continuous storage cost $u_{t,T}$, yields a convenience yield $y_{t,T}$ to its owner, in an economy where the risk-less rate is $r_{t,T}$, the no-arbitrage futures price with maturity T for that commodity will be, at time t:

$$F_{t,T} = S_t e^{(r_{t,T}+u_{t,T}-y_{t,T}).(T-t)} = S_t e^{c_{t,T}(T-t)} \qquad [1]$$

Here, $c_{t,T} = r_{t,T} + u_{t,T} - y_{t,T}$ is the cost of carry for that commodity. Under the hypothesis of market informational efficiency, the observed futures price on the market, $F_{t,T}$, should contain all available information about $c_{t,T}$: since the temperature greatly influences the cost of carry for concentrated orange juice, Roll (1984) found that futures prices for frozen concentrated orange juice possibly contain information about temperature forecasts in excess of that contained in the forecasts published by the National Weather Service. Reasoning on how the changes in futures prices embed the changes in the cost of carry is particularly interesting, because the cost of carry typically changes with physical phenomena, such as temperature.

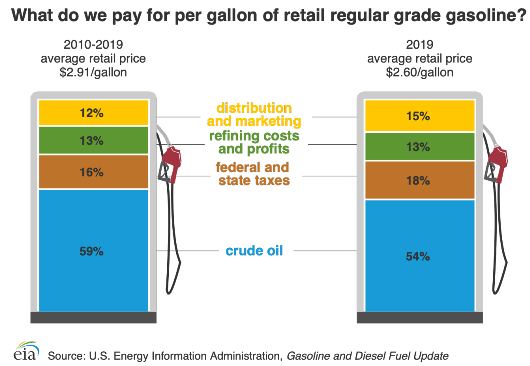

Fig. 1. Components of retail gasoline price

1. Predict Gas Prices with Oil Futures

In this section, we test the ability of crude oil futures prices to predict the price of retail gasoline. We begin by using a linear model, with parameters chosen based on economic reasoning. This serves as a benchmark against which we will measure a set of more sophisticated, non-linear models.

A. Data. The retail price for gasoline is a variable of interest for the general public. We use weekly retail gasoline prices in the United States, across all grades and all formulations, from 1993-04-05 to 2021-02-15, and the first four months of crude oil futures prices. All the data is downloaded from the EIA website. We remove the observation on 2020-04-20, where the front-month futures price went negative and had undefined log returns. The negative price was driven by the Covid-19 shock which caused demand to plummet and inventories to overwhelm storage facilities. We choose to remove this historical idiosyncrasy. We use the same test set for all models, so that the performance reports are comparable: models are evaluated on observations after 2015-12-31. The rest of the observations, from 1993-04-05 to

2015-12-31, serves for training – with, when specified, a further division between a training set and a validation set.

B. Exploratory data analysis.

B.1. Stationarity. Augmented Dickey-Fuller tests with different lags fails to reject the hypothesis that these series do not contain a unit root, as some p-values can be as large as 0.25. As these series are non-stationary, we will instead model annualized log returns of gasoline with that of crude oil futures.

B.2. Long term relationship. Augmented Engle-Granger two-step cointegration test establishes a long-term equilibrium relationship between the gasoline and crude prices with a close-to-zero p-value. In spite of that, it is imperative that we be alert to temporary decoupling due to short-term shocks or structural breaks, which are prevalent in the commodities' prices.

B.3. Volatility. The prices of crude oil futures are much more volatile than those of the retail gasoline prices. Levene tests for equal variances yield p-values close to 0. Pindyck (2001) mentions this can be due to the dampening effect of inventories in the spot market, by contrast to the futures market.

C. Benchmark: Econometric Model. We first look at the auto-correlation structure of the series. The daily returns of the futures prices are not significantly auto-correlated at any lag, but the weekly returns of retail gasoline prices are significantly auto-correlated up to lag four, while only the first partial auto-correlation of returns in gasoline prices is significant, at 0.6. This is not too surprising, given that the market of retail gasoline violates the assumptions of pure and perfect competition much more clearly than the market for crude oil: gasoline retailers exert market power to smooth prices paid by end consumers, inducing auto-correlation in retail gasoline returns. We therefore use the Newey-West variance estimator, which is robust to auto-correlation, to test our hypothesis that crude oil prices predict retail gasoline prices.

Our dependent variable is the gasoline log returns at time t, $r_{retail}(t)$. Since we want to test the ability of crude oil returns to forecast retail gasoline returns, we must include all variables which can be economically related to gasoline prices: otherwise, we might expose ourselves to an omitted variable bias. Given the first partial autocorrelation in the returns of gasoline prices is significant, we include $r_{retail}(t-1 \text{ week})$, the last available retail gasoline log returns in the dependent variables.

We test out different possible lags $r_{crude}(t - \text{k day(s)})$, and conclude that the most significant lag in the training set is $k = 1$. The contract with the strongest forecasting power is the one with maturity four months futures prices. We include its return on the trading day which precedes the day we want to forecast retail gasoline prices for. Further, we find that the "one day" period is not a strict lag: futures returns are significant on up to up to nine lags. Therefore, we add those nine lagged crude oil returns in the independent variables.

Pindyck (2001) notes that the reaction of retail gasoline prices to changes in crude oil prices may depend on the level of inventories in retail gasoline - when futures prices increase due to a temporary reason, retailers may use high inventories to ride out retail gasoline demand spikes and await reversion to normalcy. This would suggest that the coefficient on $r_{crude}(t-k\ day(s))$ should be weaker when inventories are high, reflecting a lower sensitivity to oil prices. Therefore, we create a dummy variable for high inventories, $\mathbb{1}(high\ inventories\ crude)$, defined as whether current oil inventories exceeds the twelve-month moving average, which is more robust to the structural shift in inventory level - after 2015, inventories in crude oil have been consistently higher. We also include an interaction term between the inventory level and the return on the four-month crude oil price, $r_{crude}(t-k \text{ day(s)}).\mathbb{1}(\text{high inventories crude})$.

Owyang and Vermann (2014) further mention that retail gasoline prices can respond differently in the summer and in other seasons, because higher temperatures in the summer require changes in the chemical composition of retail gasoline. Therefore, we also create a summer dummy to indicate whether the observation is in June, July, August, or September, and include $\mathbb{1}(\text{summer}).r_{\text{crude}}(t-\text{k day(s)})$.

Experimenting with these different specifications, we observe that none of the interaction effects is significant in our training data. Therefore, we choose the following baseline model:

$$r_{retail}(t) = \sum_{k=1}^{9} \alpha_k \, r_{crude}(t - \text{k trading day(s)}) + \alpha_{10} \, r_{retail}(t - 1 \text{ week})$$

In the training sample, we find that all coefficients are significantly positive and well below 1, which is conform to economic intuition: α_{10} is positive because of the auto-correlation we noticed in the returns of retail gasoline, and α_1 through α_9 exhibit the dampening effect played by inventories in spot markets, as noted by Pindyck (2001).

	Estimate (p-value)	95% Confidence Interval
α_1	0.086 (0.000)	0.038 - 0.135
α_2	0.113 (0.000)	0.073 - 0.152
α_3	0.185 (0.000)	0.141 - 0.230
α_4	0.244 (0.000)	0.204 - 0.284
α_5	0.276 (0.000)	0.224 - 0.327
α_6	0.168 (0.000)	0.126 - 0.211
α_7	0.161 (0.000)	0.122 - 0.199
α_8	0.086 (0.000)	0.038 - 0.133
α_9	0.048 (0.019)	0.008 - 0.089
α_{10}	0.511 (0.000)	0.394 - 0.627

Table 1. In-Sample Coefficient Estimates for Baseline

To assess the stability of the relationship over different horizons, we run rolling regressions by windows of 200 weeks, and look at the estimated values of the coefficients for each window, by date when each window begins (Figure 2). In spite of fluctuations in the coefficients with time, notably following financial crises of 2001 and 2008, the relationship is remarkably stable: the order of magnitude of the coefficients does not change much. Therefore, we execute an expanding regression, using all the information up to the day before the day we want to forecast the return in retail oil prices for.

In-sample, the performance of our baseline is excellent. Given the high number of variables it includes, it seems to partly over-fit the training set. However, even out of sample, the baseline explains an important portion of the variance.

	1000*MAE	1000*MSE	Explained Variance (%)
Train	8.38	0.14	59.91
Test	8.78	0.14	43.01

Table 2. Model Results for the Baseline

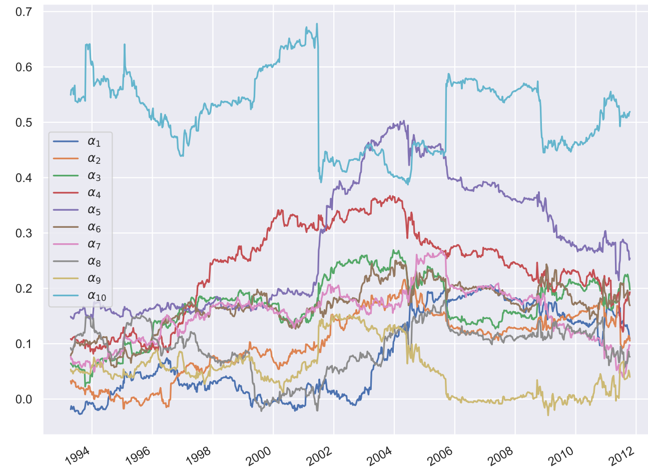

Fig. 2. Coefficients of rolling regressions

This result, coupled with our formal test for the significance of the coefficients, confirms that crude oil futures prices have some forecasting power for the retail prices of gasoline. We will now look into more complex models and try to beat the baseline.

D. Vector Error Correction Model (VECM). Graphically, we see gasoline and oil prices tend to move together. To formally validate this idea, we use VECM to test the co-integrating effect, and if the term structure in the oil futures market help with the prediction. We model weekly gas and oil prices as

$$X = \begin{bmatrix} \text{Retail gas price} \\ \text{Front month oil futures price} \end{bmatrix}$$

$$Y = \begin{bmatrix} \text{2nd month oil futures basis} \\ \text{3rd month oil futures basis} \\ \text{4th month oil futures basis} \end{bmatrix}$$

$$\Delta X_t = \alpha\beta' X_{t-1} + \Gamma \Delta X_{t-1} + \phi' Y_{t-1} + \varepsilon_t$$

where the n-th month oil futures basis is defined as the difference between the price of such futures and the front month futures, β is the vector representing co-integrating relations. We do not include the price level of back-month futures as they form a much stronger co-integrating relationship with the front-month contract, which may pollute the more subtle relationship between gas and oil prices. The proposed model is valid as the basis is found to be statistically stationary.

Table 3 displays the result of VECM, where the signs of the co-integrating vector are aligned with economic intuition. When oil futures price is high, the co-integrated vector is lowered, bringing up the retail gas price.

	Retail gas price	1st oil futures price
Co-integration Vector β	1	-0.038**(0.000)
The p-value under Augmented Dickey-Fuller test is less than 0.001.		

Regression Result

	Dependent Variables (Change)	
Regressor	Retail gas price	Front-month oil futures price
Co-integration Vector	-0.006**(0.033)	0.370**(0.042)
2nd month oil futures basis	0.019*(0.090)	-3.515***(0.000)
3rd month oil futures basis	-0.003(0.885)	4.589***(0.000)
4th month oil futures basis	-0.005(0.564)	-2.104***(0.001)
Lagged (Change)		
Retail gas price	0.473***(0.000)	8.905***(0.000)
Front-month oil futures price	0.006***(0.000)	-0.031(0.297)

* p<0.10, ** p<0.05, *** p<0.01

	1000.MAE	1000.MSE	Explained Variance (%)
VECM (rolling)	190.81	386.24	49.05

Note:
VECM predicts price instead of returns.
We fit a new VECM model as time progresses each day.

Table 3. Results of VECM. The signs in relation to the co-integration vector have meaningful economic explanation.

We also experimented with the term structure of convenience yield, which we extracted from the first four oil futures in the light of equation 1. Assuming constant oil storage cost and bootstrapping the LIBOR curve, we obtained the level and slope of the convenience yield curve, which should be reflective of expectation about energy demand and oil inventory. The latter helps energy producer to meet fluctuation in demand without disrupting production schedule. Reasonable as it may be to believe these effects would impact retail gas prices, we find that neither the convenience yield nor its slope has significant effects. This is probably because gasoline supply chain is so streamlined and competitive that only the immediate oil supply impacts retail pricing.

E. Deep Learning Model. VAR and VECM succeeded in explaining some variance in the gasoline returns while being economically intuitive and parsimonious. However, in order to harvest non-linear information - unstable interaction between inventory (actual and predicted), futures term structure, energy demand outlook -, fitting a low bias deep neural network to them may uncover more insights.

We explored the LSTM, CNN, CNN-LSTM architectures Livieris et al. (2020), Pintelas et al. (2020) and found the LSTM performed the best out-of-sample. Deep-learning models capture non-linearity with multifarious types of activation functions like ReLU, Sigmoid, tanh. Feedback connections, memory cells and internal gates in LSTM empower it to model time series and long-term temporal structures without vanishing gradient problems.

Our train, cross-validation and test set run, respectively, from 1993 to 2012, from 2013 to 2015, and from 2016 onwards, except that we skip a day between the periods to prevent information leakage. To ensure our training data are range-bounded, we scale our input relative to the min and max value in the training set. We impute missing data with the median and and clip the outliers. We then club together the past four time steps / four weeks (look-back period) of data points. This results in a tensor of the shape: # samples x # time steps x # features.

The features we use include the lagged log returns of gasoline spot, crude oil, crude oil futures contracts, gasoline inventory surprise and refinery margins. The gasoline inventory surprise is computed as the standardized actual inventories in the market minus the forecasted inventory. The refinery margins are estimated as the difference between the cost at which refineries buy crude oil, and the price charged to the retail agents, divided by the sale price.

The default LSTM model architecture includes one 'LSTM' layer of 5 units with 'reLU' activation, followed by a fully connected 'Dense' layer with single output for one step ahead forecast of retail gasoline log-returns. To avoid over-fitting, we use elastic net regularization and dropout in the LSTM layer. Our loss is the 'mean absolute error'(MAE) rather than 'mean squared error' (MSE): the MSE becomes very small for log returns, which would make training the model very challenging. ADAM with batch training (batch size = 200) is used to minimize the loss.

We then evaluate the model on the train, CV and test sets.

We begin with testing a model using only lagged gasoline log returns. We subsequently add log returns of the lagged crude oil first futures contract to see the value add: the explained variance jumps from around 8.5 percent to 17 percent. We see maximum improvement in the model performance across

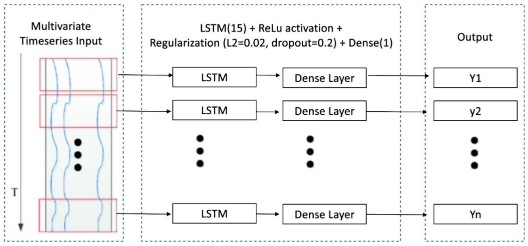

Fig. 3. LSTM Architecture: Grid Search Optimal

train and CV sets when considering refinery margins along with log returns of lagged gas spot, crude oil spot and the first four futures contracts. 4 summarizes the model performance across the data sets.

Finally, we tune the hyper-parameters: the learning rate, the number of units in the LSTM model, the activation functions, dropout, and regularization hyper-parameters. We employ Bayesian optimization based on grid-search. Unlike other grid search methods, Bayesian optimization monitors past evaluation results and smartly chooses the next best parameters, thus calling the objective function less frequently, which results in a better model in fewer iterations compared with random or linear grid search. The optimal model architecture is shown in 3.

F. Regime switching model. While Deep Learning models have the ability to capture complex and time-varying relationships, their output may be sub-optimal if the "true" pricing relation involves distinct regime-specific models. Thus, we also wanted to test whether regime-switching techniques could provide a better empirical model of retail gas price dynamics. The most primitive usage of the regime models was put into practice by the seminal work of Hamilton (1989).

PCA of crude future prices' log-returns. We began with a principal component analysis on the crude futures log-returns term structure using the first 4 future contracts. The first two components explain more than 99% of the variance. The objective of the PCA is two fold: first, it encapsulates the variance originating from interest rates and the convenience yield term structure. Second, it allows to specify the model parsimoniously.

The first principal component (CFPC1), which explains the majority of the variance in the crude future returns, is selected to point in the direction where it has a positive correlation to the front-month crude future. This is done so that the actual sensitivity of the gasoline prices to front-month futures and sensitivities for the gasoline prices to the PC1 from the model are comparable. Generally, in regime switching models for econometric variables, it is a reasonable choice to design a model with two regimes. The choice of two states can be associated with periods of high and low values of volatility, which is intuitively appealing. We assume different means and standard deviations in each regime to avoid any misspecification. However, it is imperative to note that regimes are not labelled before any modeling happens: they are the output of a purely quantitative modeling exercise.

Table 4. Deep Learning (LSTM) model evaluation metrics

Gas Spot and Lags only

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	13.42	0.38	8.62
CV	13.65	0.36	8.88
Test	12.89	0.29	8.99

Crude Spot, First Futures Contract and gas Spot

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	13.36	0.37	16.44
CV	13.61	0.36	17.26
Test	12.7	0.29	18.41

Crude Spot, First 4 futures contracts and Gas Spot

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	12.65	0.33	16.97
CV	13.6	0.35	16.62
Test	11.66	0.25	23.25

Crude Spot, First 4 futures contracts, Gas Inventories and Gas Spot

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	13.15	0.36	11.19
CV	13.03	0.32	22.59
Test	12.32	0.28	23.07

Crude Spot, First 4 futures contracts, Refinery Margins and Gas Spot

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	12.42	0.32	19.5
CV	13.22	0.33	20.02
Test	11.53	0.24	24.94

Crude Spot, First 4 futures contracts, Refinery Margins and Gas Spot: Grid Search

	1000*MAE	1000*MSE	Explained_Variance (%)
Train	13.21	0.33	39.71
CV	12.63	0.31	38.89
Test	12.34	0.26	41.32

We tested a large selection of models, comprising various lags in gasoline and principal components of crude futures. We found the model below to be most appropriate.

$$r_{GP,t} = \gamma + \alpha_1 r_{GP,t-1} + \alpha_2 r_{GP,t-2} + \alpha_3 r_{CFPC1,t-1} + \alpha_4 r_{CFPC2,t-1} + \alpha_5 ECT_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma_S^2)$, $S \in \{0,1\}$ represents the variance for a regime, $r_{GP,t}$ refers to the gasoline price returns, $r_{CFPC1,t}$ and $r_{CFPC2,t}$ refer to the first two principal components obtained from the PCA of first 4 crude futures returns, and ECT corresponds to the residuals from the co-integration relationship between Gasoline prices and front-month future contract and intends to capture the cumulative effect of long-term relationship on short-term variability.

The best fit model selection is done based on AIC. In absence of a large difference between the model's log-likelihood, the selection is based on the Regime Classification measure introduced by Ang and Bekaert (2002) in their study of regime study for interest rates.

$$RCM = 100 * k^2 * \frac{\sum_{j=1}^{n} \prod_{i=1}^{k} p_{i,j}}{n}$$

where $p_{i,j}$ represents the smoothed probabilities of an observation j in state i and, therefore, $\sum_{i=1}^{k} p_{i,j} = 1$. The lower the value of RCM, the model calibrates to more distinctive regimes across time. The result and error metrics for the model are in Tables 5 and 6.

Table 5. Model results for gasoline regime switching model. q represents the transition probability

	Regime=0	Regime=1
γ	-0.0008 (0.0043)	-0.0008 (0.3125)
α_1	0.3506 (0.0)	0.3814 (0.0)
α_2	0.2307 (0.0)	0.1476 (0.0)
α_3	0.0017 (0.0)	0.0036 (0.0)
α_4	-0.0014 (0.2556)	-0.0056 (0.0158)
α_5	-0.0123 (0.0046)	-0.0422 (0.0)
σ^2	0.0 (0.0)	0.0003 (0.0)
$q_{i,i}$	0.957	0.976

Table 6. Error metrics for gasoline regime switching model

	1000*MAE	1000*MSE	Explained variance(%)
Train	9.06	0.19	48.96
Test	9.44	0.19	45.72

The transition of the states is depicted in Figure 4. We observe some change of regime after year 2000. The period before 2000 is mostly dominated by the low volatility.

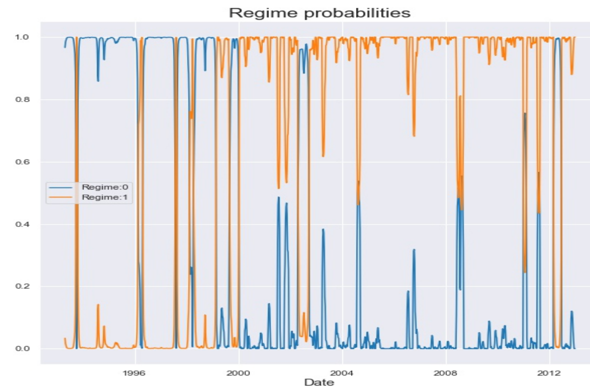

Fig. 4. Gasoline price transition probabilities

G. Asymmetric relationship. Finally, we extend the relationship obtained from the previous sections to establish whether an asymmetric relationship exists, by introducing the asymmetric terms (represented by the superscript $+$) in our model specification.

$$r_{GP,t} = \gamma + \alpha_1 r_{GP,t-1} + \alpha_2 r_{GP,t-2} + \alpha_3 r_{CFPC1,t-1} + \alpha_4 r_{CFPC2,t-1} + \alpha_5 r^+_{GP,t-1} + \alpha_6 r^+_{GP,t-2} + \alpha_7 r^+_{CFPC1,t-1} + \alpha_8 r^+_{CFPC2,t-1} + \alpha_9 ECT_{t-1} + \varepsilon_t$$

Here $r^+_{GP,t} = r_{GP,t-1} * \mathbb{1}_{r_{GP,t-1}} > 0$, and $r^+_{CFPC1,t-1}$ and $r^+_{CFPC2,t-1}$ are similarly defined. The sensitivity of the gasoline returns to crude oil returns is measured as $\alpha_3 + \alpha_7$ for increases in crude prices and α_3 for decreases in crude prices. If no asymmetric relationship exists, α_7 should be insignificant.

In the low volatility regime 0, a 1% increase in $CFPC1$ corresponds to a 0.0024% change in the gasoline prices, whereas 1% decrease corresponds to 0.0013% decrease. Consequently, it can be concluded that in the low volatility phase, retailers are quick to adjust price upwards when the crude price increases, but they do not decrease proportionally in case of drop in oil prices.

In the high volatility regime 1, a 1% increase in $CFPC1$ corresponds to a 0.0033% change in the GP whereas 1% decrease corresponds to 0.0037% decrease. Therefore, unlike the difference in sensitivities in the low vol regime, there is not a significant difference in sensitivity to oil prices in the high volatility regime. The asymmetric relationship seems more prevalent in the low volatility regime than in the high volatility regime. These results are in line with intuition: when volatility is high, retailers could be more concerned about the demand-supply dynamics and have a myopic view of the market, which leads to changing gasoline prices with the changing crude prices.

Table 7. Model results for gasoline regime switching asymmetric model

	Regime=0	Regime=1
γ	-0.0014 (0.0289)	0.0047 (0.0003)
α_1	0.4194 (0.0)	0.6616 (0.0)
α_2	0.1586 (0.0691)	0.2239 (0.0035)
α_3	0.0013 (0.0001)	0.0037 (0.0)
α_4	-0.0046 (0.0071)	-0.0113 (0.0103)
α_5	-0.2014 (0.1484)	-0.49 (0.0)
α_6	0.1058 (0.3977)	-0.2081 (0.0406)
α_7	0.0011 (0.0766)	-0.0004 (0.6749)
α_8	0.0034 (0.1937)	0.0108 (0.1049)
α_9	-0.0129 (0.0054)	-0.0365 (0.0)
σ^2	0.0 (0.0)	0.0003 (0.0)
$q_{i,i}$	0.966	0.98

Table 8. Error metrics for gasoline regime switching asymmetric model

	1000*MAE	1000*MSE	Explained variance(%)
Train	8.85	0.18	52.95
Test	9.09	0.17	50.20

H. Additional factors. Figure 1 shows that, although crude oil is the primary component of the price of retail gasoline, other factors matter. Specifically, costs and profits for the refiners and distributors can vary widely. Below, we outline additional factors that could impact the prices, and account for unexplained variance.

H.1. Policy risk. Gasoline prices increase when the demand is greater than the supply of the commodity. The supply of crude is majorly controlled by the policies from the OPEC nations, which is difficult to forecast.

H.2. Complementary and substitution goods. While the automobiles are a major consumers of gasoline, the emergence of electric and other cleaner fuels and taxes levied on the auto industry influence the demand for gasoline. Shift in the demand on these substitution goods might change the price of gasoline.

2. Predict Atmospheric CO_2 Content

Carbon dioxide is one of the main drivers of climate change, which threatens the sustainability of our economy, and represents a new market risk factor. Key factors affecting atmospheric carbon dioxide concentration include natural causes like volcano eruptions, biological life-cycles, seasonal temperature variations, and human activities like deforestation and, most importantly, combustion of fossil fuels. The demand of energy is closely connected to the economy and therefore we investigate if the financial markets, particularly those for the most important fossil fuel - oil - provide any predictive power for atmospheric carbon dioxide content.

A. Seasonal Variation. As seen in the top panel of Figure 5, atmospheric carbon dioxide exhibits a long-term uptrend and strong seasonal patterns. CO_2 concentration levels usually come down during the spring and summer seasons due to the increased photosynthesis from the plants and absorption from the ocean surface.

To break down this time series into different additive components, Robert et al. (1990) proposes using locally estimated scatter-plot smoothing (LOESS) to extract an estimate of the long-term trend on the first pass, and LOESS again to extract the seasonal components. Figure 5 and Table 9 show the components of the decomposition and the summary of the results. The left panel of figure 6 shows that the residuals from the STL treatment has largely stripped away cyclicity of carbon dioxide level, and we should be able to apply regular econometric techniques. The right panel depicts the residuals.

	Mean	Std	ADF p-value	DW Stat
CO2 concentration	366.552	23.547	0.973	0.000
CO2 trend	363.077	20.911	1.000	0.000
CO2 season	0.012	2.197	0.000	0.065
CO2 residuals	-0.009	0.566	0.000	1.776
CO2 trend (1st diff)	0.034	0.010	0.001	0.000

Table 9. Summary of trend, seasonality and residuals of atmospheric carbon dioxide concentration from May 1974 to November 2019. Source: Thoning et al. (2020)

B. Trends and Residuals Explained by Financial Markets. Oil prices have long been used as a proxy for global economic activity and demand for energy,

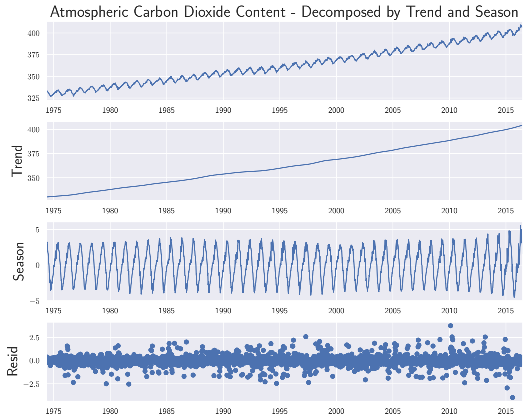

Fig. 5. Breakdown of atmospheric carbon dioxide concentration into long-term trend, seasonal component and residuals using robust STL decomposition. Robert et al. (1990)

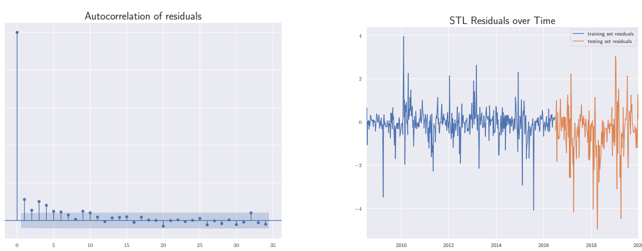

Fig. 6. Left: ACF of STL residuals; Right: Residuals from STL performed on training window (pre-2016) and extrapolated into testing window (post-2016). The cyclicity of CO_2 levels have been removed by STL.

so it is a reasonable conjecture that oil prices may predict carbon dioxide emission. For example, as energy demand increases, more oil is consumed and the price of oil will rise. Ambiguously, we could also argue that lower energy prices encourage energy consumption. In addition to energy prices, we are also interested in the prices of carbon credits, which are transferable permits giving the right to emit a certain amount of carbon dioxide. An example of carbon credits is the European Union Emission Trading Scheme (EU ETS). As energy demand rises, energy producers emit more carbon dioxide, driving carbon emission allowance prices higher. The intention of regulators is that this higher price will eventually deter emissions. In our view, it remains an empirical question as to whether carbon credit prices merely act as a proxy for economic activity and thus positively predict future CO2 or whether higher carbon credits deter emissions and result in lower CO2.

For the purpose of this econometric analysis, we assume the seasonal patterns observed in carbon dioxide content is attributable to the natural seasonal cycle of the Earth, and the effects of human activity are manifested only in the long-term trend and short-term residuals around this trend. We therefore assess to what extent the prices of oil futures and carbon emission allowances explain the variation in the long-term trend and seasonal residuals.

First, we fit a linear model, and then proceed to use Deep Learning Models to capture both the linear and the non-linear relationships. We use August 2008 - June 2016 as a training set and July 2016-November 2019 as test period. EU ETS carbon credits commenced trading in 2008.

Model for CO_2 residuals

Regressor	Coefficient	Std. error	t-stat(p-value)
Intercept	-0.053	0.033	-1.618(0.106)
Lagged:			
CO_2 residuals	0.103	0.049	2.103^{**}(0.035)
CO_2 futures return	-0.799	0.375	-2.129^{**}(0.033)
Oil futures return:			
Front month	-1.110	0.837	-1.326(0.185)
Next month (basis)	-8.021	5.969	-0.502(0.615)
Third month (basis)	34.263	9.442	0.869(0.385)
Fourth month (basis)	-25.856	6.486	-0.976(0.329)

Model for CO_2 trend change (1st diff)

Regressor	Coefficient	Std. error	t-stat (p-value)
Intercept	-0.000	0.000	-0.909 (0.363)
Lagged:			
CO_2 trend (1st diff)	1.004	0.002	511.599^{***} (0.000)
CO_2 futures return	0.000	0.000	0.371 (0.711)
Oil futures return:			
Front month	-0.001	0.000	-1.132 (0.258)
Next month (basis)	0.004	0.009	0.425 (0.671)
Third month (basis)	-0.007	0.021	-0.331 (0.741)
Fourth month (basis)	0.004	0.014	0.249 (0.804)

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 10. Summary of models explaining long-term trend and seasonal residuals

Table 10 displays the result of fitting a VAR(1) for CO_2 residuals to each of the variables in index. The back-month oil futures basis are taken as the return of such futures less that of the front-month contract. We observe that the most statistically significant coefficients are that of lagged CO_2 residuals and CO_2 futures, while the coefficients of the four oil futures are insignificant and largely offsetting. Even when we only include the front-month oil futures in a VAR model, it fails to produce a significant coefficient. The positive CO_2 residual coefficient is elucidated in Figure 6, which may be explained by stickiness of factors affecting carbon dioxide levels.

For instance, an extraordinarily busy period for factories is likely to be followed by another busy period, resulting in sustained anomaly in the STL residual even adjusted for seasonality. Oil prices, which we view as a proxy for economic activity, do not add explanatory power above what is contained in residual time series. As for the negative coefficient of CO_2 futures, it suggests that an increase in prices is connected to a reduction in atmospheric carbon dioxide content. It appears to be in line with the rationale of implementing carbon cap-and-trade mechanisms, deterring further carbon emissions when the demand to emit is high. We do not observe statistically meaningful results from regressing the CO_2 trend change against other variables. Even though the lagged residuals and carbon credit futures returns may have some significant effects over the residuals, their explanatory power of the carbon dioxide level is dwarfed by the long-term trend and seasonality pattern. Table 11 compares the out-sample performance of various models, and it suggests that CO_2 futures returns does not improve explained variance relative to a simple STL. This is not surprising as these variables only partially predict the residual, which itself is a tiny component of overall atmospheric carbon dioxide. Figure 7 depicts the prediction of the VAR(1) model and the true historical values. We conclude that the most important factor, at least for linear models, is the trend and seasonality of carbon dioxide levels, followed by its lag. Carbon credits and oil futures prices can, at best, marginally help in predicting future carbon dioxide levels.

	1000x MAE	1000x MSE	Explained Variance (%)
STL sesonality and trend only	845.0	1492.0	88.803
STL+AR(1) with lagged:			
CO2 levels	845.1	1492.2	88.803
CO2 levels + oil futures	844.6	1490.8	88.805
CO2 levels + CO2 futures	863.7	1541.7	88.821
CO2 levels + oil & CO2 futures	1265.8	2613.3	88.004

Table 11. Out-of-sample performance of STL with various models explaining the residuals. Using lagged CO_2 levels and CO_2 futures return give the best result.

C. Deep Learning Model. We build upon the seasonal and trend decomposition of carbon dioxide using LOESS / robust STL technique. We focus on using the LSTM model for capturing the CO_2 residual patterns while considering the following explanatory variables: lagged CO_2 residuals, log returns of the carbon credits (CO_2 futures) and crude oil front four futures contracts.

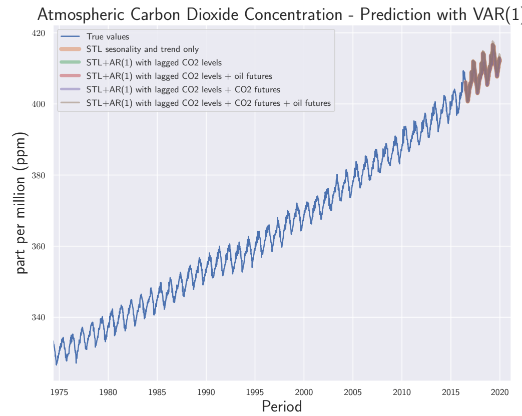

Fig. 7. Predictions of various VAR(1) models and true historical carbon dioxide concentration, where the seasonal component is projected with STL decomposition. Robert et al. (1990)

We use the same LSTM architecture from the gasoline model, but this time, we adopt a look-back window of 52 weeks in order to capture any leftover seasonality in the residuals and/or the explanatory variables.

We pre-process the explanatory variables by clipping the outliers beyond 0.5 percentile in the tails of the distributions and filling the missing values with the median of the past observed values. We consider three variants in our analysis: predicting CO_2 residuals using (1) the lagged CO_2 residuals only, (2) using lagged CO_2 residuals and front four crude oil futures, and (3) lagged CO_2 residuals and carbon credit futures (CO_2 futures).

We find that lagged CO_2 residuals help explain about 2.23% of the CO_2 residuals in the test set, which comprises of latest 20% of the total data. When we add the first four futures contracts, the explained variance rises an insignificant amount to 2.56%. However, the explained variance for third variant, which adds the CO_2 futures to the residuals, increases to 8.52%. Therefore, the output from both the linear and LSTMs model suggest CO_2 futures have some ability to predict CO_2 residuals. As a final step in the LTSM modelling, we conduct Bayesian optimization-based grid search to perform hyperparameter tuning. The results are summarized in Table 12.

Looking at the predictions distribution from our best performing model (Figure 8), it appears CO_2 residuals predicted by lagged CO_2 residuals and CO_2 futures show a positive mean and fat tails. This result is intuitive as increasing economic activity over time should lead to more positive surprises from the trend and seasonality in CO_2 concentration.

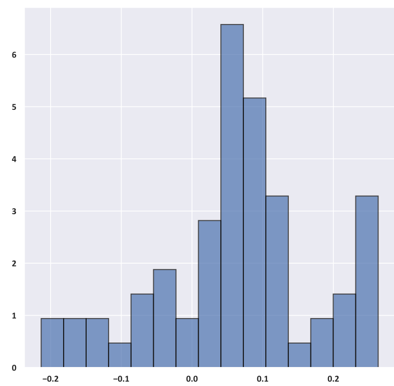

Fig. 8. Distribution Of Predicted CO2 Residuals

Table 12. LSTM Model error metrics for CO_2 Residuals

Lagged CO_2 Residuals only			
	1000*MAE	1000*MSE	Explained_Variance (%)
Train	260.8	124.83	6.55
Test	492.5	566.53	2.23

Lagged CO_2 Residuals + First 4 Crude futures			
	1000*MAE	1000*MSE	Explained_Variance (%)
Train	198.9	85.65	8.82
Test	597.69	640.31	2.56

Lagged CO_2 Residuals + CO2 futures			
	1000*MAE	1000*MSE	Explained_Variance (%)
Train	192.64	79.84	14.45
Test	553.4	579.52	8.52

Lagged CO_2 Residuals + CO2 futures: Grid Search Optimized			
	1000*MAE	1000*MSE	Explained_Variance (%)
Train	287.06	116.35	18.94
Test	567.14	559.05	11.9

3. Conclusion

In this report, we examined if and to what extent crude futures prices can predict real-life phenomena. We found that gas prices at the gas station can be effectively predicted with lagged oil futures prices. We examined the relation with a wide range of models - linear regression, VECM and VAR models, regime-switching and deep learning models. We also tested the relevance of several connected economic variables, including oil and gas inventory, refinery margins and asymmetric effects of energy prices.

We then assessed to what extent financial markets encode information about future carbon dioxide concentration. We examined both oil prices and prices for EU carbon credits. Carbon credits appeared to have more predictive power than oil prices, but the link was restricted to just residual changes in CO_2 after removing seasonal and trend components. Our conclusion thus conforms with economic intuition. In markets like gasoline, where oil is a primary input to a straightforward production process, oil prices contain some information about future prices. In the situation of carbon dioxide, where the chain of causality is exceptionally complex, oil prices have no meaningful predictive value, although carbon futures prices have some limited predicted power for movements in residuals around a long-term, seasonally adjusted trend.

References

Ang, A. and Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business Economic Statistics*, 20(2):163–182.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Livieris, I., Stavroyiannis, S., Pintelas, E., and Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. *Neural Computing and Applications*.

Owyang, M. and Vermann, E. K. (2014). Rockets and Feathers - Why Don't Gasoline Prices Always Move in Sync with Oil Prices? *The Regional Economist*, (October).

Pindyck, R. S. (2001). The Dynamics of Commodity Spot and Futures Markets: A Primer. *The Energy Journal*, pages 1–29.

Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., and Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. In Maglogiannis, I., Iliadis, L., and Pimenidis, E., editors, *Artificial Intelligence Applications and Innovations*, pages 99–110, Cham. Springer International Publishing.

Robert, C., William, C., and Irma, T. (1990). STL: A seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics*, 6(1):3–73.

Roll, R. (1984). Orange Juice and Weather. *The American Economic Review*, 74(DECEMBER):861–880.

Thoning, K., Crotwell, A., and Mund, J. (2020). Atmospheric Carbon Dioxide Dry Air Mole Fractions from continuous measurements at Mauna Loa, Hawaii, Barrow, Alaska, American Samoa and South Pole.