The Dynamics of Storage Costs*

Andrei Stancu[†], Lazaros Symeonidis[‡], Chardin Wese Simen[§], and Lei Zhao[¶]

December 18, 2022

Abstract

We document that the monthly storage cost of oil averages 0.50% of the spot price and varies over time. We decompose the *basis*, defined as the ratio of the spread between the futures and spot prices over the spot price, into the storage cost (*scc*) and the adjusted convenience yield (*acyc*) channels. The *scc* dominates the mean of the *basis* and accounts for nearly half of its variations. We show that the *scc* predicts future inventory growth and is the main conduit through which the predictive power of the *basis* for oil spot returns arises.

JEL classification: G13, G14, G17

Keywords: Basis, Futures, Inventory, Predictability, Storage Costs

*We are immensely grateful to Perry Cooper for very detailed conversations about the inner workings of the LOOP storage futures market. We also thank the staff of S&P Global Platts for very insightful discussions about the market for storage. We thank Alex Kostakis, Bartosz Gebka, Brian Wright, Charlie Cai, Darren Duxbury, Eugenio Bobenrieth, Kristoffer Glover (Discussant), Lazlo Koczy (Discussant), Lucio Sarno, Marceau Tchouaga, Marcel Prokopczuk, Massimo Guidolin, Paul Karehnke, Tom Lord, Xiaoxia Ye, and participants at the 2021 meetings of the Commodity and Energy Markets Association, the 2021 ACFR Derivative Markets Conference, the Inaugural Conference in Honor of Lucio Sarno, the 5^{th} Annual JP Morgan Commodity Centre International Commodities Symposium, Leibniz University Hannover, Newcastle University, the University of Bath, and the University of East Anglia. This research is supported by funding from the British Academy Small Research Grant, ESCP Europe, and the University of Liverpool. Contact: andrei.stancu@newcastle.ac.uk. (A. Stancu), l.symeonidis@essex.ac.uk (L. Symeonidis), C.Wese-Simen@liverpool.ac.uk (C. Wese Simen), and lzhao@escp.eu (L. Zhao).

[†]Newcastle University Business School, Newcastle University, Newcastle Upon Tyne, NE1 4SE, United Kingdom.

[‡]Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom.

[§]Management School, University of Liverpool, Liverpool, L69 7ZH, UK.

[¶]Finance Department, ESCP Business School, Paris Campus, 79 Avenue de la Republique, 75543 Paris Cedex, and Labex ReFi, France.

I Introduction

Inventories play a key role in the literature on commodity markets. They are at the heart of theoretical frameworks such as the theory of storage (Kaldor, 1939; Working, 1949; Brennan, 1958). Empirical studies are increasingly interested in inventory data too. For instance, Armstrong et al. (2021) explore the impact of inventory news on market activity, while Ederington et al. (2021) analyze the impact of cash-and-carry arbitrage on inventory levels. Despite the central role that inventories play in the literature on commodity markets, it is rather surprising that we know so little about the cost of storing these inventories. Therefore, several interesting questions arise: What is the average cost of storage? Is this cost constant or time-varying? What are the implications of the time-series properties of the storage cost for the dynamics, economic interpretation, and predictive power of the *basis*, defined as the ratio of the spread between the futures and spot prices over the spot price (Boons and Prado, 2019)?

Answering these questions is important because several studies, including the recent papers of Gu et al. (2020) and Ederington et al. (2021), assume that the cost of storage is small and/or exhibits very little time-series variation.[1] Despite the widespread nature of these assumptions in the academic literature, we are not aware of any empirical test of these hypotheses.[2] The dearth of research on the cost of storage is likely due to the fact that information about the cost of storage is often proprietary and not readily available

[1]This assumption may help explain why the literature routinely interprets the spread between the futures and spot prices as the net convenience yield rather than net storage cost. See Casassus and Collin-Dufresne (2005) for example.

[2]If anything, anecdotal evidence points to the opposite direction. Indeed, conversations with S&P Global Platts reveal that the price of onshore storage varies between 10 and 50 cents per barrel per month.

to academics.

We use a novel dataset of the Louisiana Offshore Oil Port (LOOP) sour crude oil storage futures contract (SFC) to shed light on the dynamics of storage costs. This contract, which started trading in March 2015, gives the holder the right to store 1,000 barrels of crude oil at the Clovelly Hub, one of the largest private storage facilities in the U.S., during the delivery month. The SFC is well-suited for our analysis since it makes the pricing of storage transparent to all market participants. We document several novel findings. First, the cost of storing oil for 1 month corresponds to 0.50% of the spot price of oil on average. This proportional storage cost exhibits considerable fluctuations as evidenced by the (monthly) volatility estimate of 0.89 %. These time-variations are economically meaningful too. For instance, the average proportional storage cost is higher during contango periods (0.74%), when the incentive to store is strong, than during backwardation periods (0.13%), when there is little incentive to store oil.

Second, we complement our data of SFC with the dataset of the LOOP Gulf Coast sour crude oil futures contract (GCOFC), which is also physically settled at the Clovelly Hub. Exploiting the insight that the SFC and the GCOFC are both physically settled at the same location, we propose a decomposition of the basis into (i) the storage cost channel (*scc*) and (ii) the adjusted convenience yield channel (*acyc*). We find that the two channels are negatively correlated. Furthermore, the decomposition reveals that the *scc* dominates the level of the *basis* and accounts for nearly half of its variations. This set of results is particularly strong during contango periods and clearly challenges the standard assumptions in the literature that the storage cost is small (Ederington et al.,

2021) and/or constant (Gu et al., 2020).

Third, we analyze the information content of the SFC. If the SFC is informative about the cost of storing oil over the delivery month, it is interesting to analyze its predictive power for next-month's inventory level. We regress the 1-month inventory growth rate on a constant and the lagged growth rate of the *scc*. We find a positive and statistically significant slope estimate (0.020, t-stat=2.802). The adjusted R^2 of 7.07% further confirms the information content of the SFC. We augment the model with the lagged *acyc* and various control variables used in the literature, e.g., Ederington et al. (2021), and obtain very similar results. Collectively, these results suggest that market participants generally pay more for the SFC when they plan to have more inventories.

Fourth, we document that the basis predicts the spot returns of the LOOP Gulf Coast crude oil. Given our decomposition of the basis, we evaluate the relative contribution of each of the two components to the predictive power of the basis. Our analysis reveals that the *scc* predicts the spot return with a positive and statistically significant loading (3.800, t-stat=3.334). The economic message is simple and intuitive: market participants pay more for storage when they anticipate higher future spot returns. In contrast, the slope associated with the *acyc* is not statistically significant. Taken together, this set of results suggests that the *scc* is the main channel through which the predictability of the *basis* arises. We are not aware of any study documenting this result.

We conduct several additional analyses. We begin by exploring the sensitivity of our inventory forecasting regressions to potential information leakage ahead of the publication of the inventory data. Next, we conduct a bootstrap experiment to account for the fact that our sample period is relatively short, making the statistical inference potentially treacherous. The bootstrap analysis confirms the robustness of our results. The cost-of-carry formula requires us to take a stance on the proxy for the interest rate. To this end, we assess the sensitivity of our findings to the interest rate proxy and find that it does not materially affect our results. Additionally, we consider alternative definitions of oil market conditions and find no discernible effect on our main findings. Furthermore, we investigate the contribution of the storage cost channel to the convexity measure of Gu et al. (2020). We find that the cost of storage plays an important role in the convexity measure. Finally, we document the predictive power of the *scc* for the spot return of the WTI crude oil as well as the return on the exchange traded fund (ETF) tracking companies that are active in the transportation and storage segment of the energy value chain.

Our work is directly related to the literature on the cost of storage. Using warehouse data for June 1984, Fama and French (1987) document large cross-sectional differences in the monthly storage cost with estimates varying between 0.01% and 2.67% of the corresponding spot price. Ross (1997) estimates that the 1-month cost of storage represents 1.67% of the spot price of oil.[3] Routledge et al. (2000) assume that the 1-month storage cost is equal to 0.25% of the oil spot price. In a parametric setting, Byun (2017) estimates that, when the price of the WTI crude oil reached $99.21 per barrel, the marginal cost of storing oil for 1 month was 0.91% of the spot price. Baker (2021) assumes that the cost of storing oil for 3 months represents about 3 % of the spot price of oil. Chincar-

[3] To be specific, the author reports that the cost of storing a barrel of oil for 1 month is $0.25. The price of oil at the time was around $15.

ini (2020) documents large variations in the storage costs estimates contained in analyst reports. The large variations in the estimates of the extant literature might arise from the fact that the storage cost estimates (i) appear to be selected on an ad-hoc manner in theoretical papers, (ii) depend on the modelling assumptions, (iii) vary with the type of storage, e.g., salt cavern, railroads, or onboard tankers at sea, (iv) change with the length of the lease contract, and (v) are based on proprietary data sources.[4] Different from these studies, we use data from the storage futures market, which make the pricing of storage transparent for each delivery month. We document a novel set of stylized facts regarding the cost of storage that can be used to further discipline and evaluate theoretical models.

Our paper also contributes to the commodity literature that assumes that storage costs are relatively stable. This literature dates back at least to the work of Brennan (1958). The author surveys the price of cold-storage for some dairy and agricultural commodities and notes that the storage cost is generally the same from one month to the next.[5] Gu et al. (2020) point out that the cost of storage may vary across commodity markets but not in the time-series dimension. To the best of our knowledge, we are the first to empirically study the dynamics of the cost of storage. We quantify the contribution of the *scc* to the dynamics of the *basis*. Our estimates suggest that the *scc* accounts for close to half of the variations in the basis over our sample period. At a minimum, our results caution against the usual interpretation that the basis is mostly informative about the convenience yield.

[4]Using a proprietary dataset from Plains All, the largest storage operator at Cushing, Ederington et al. (2021) document that the mean difference between the annual maximum and minimum storage costs is \$0.175. Unfortunately, this information is difficult to interpret in a meaningful manner as the authors do not provide statistics related to the average monthly storage cost.

[5]Specifically, Brennan (1958) obtains the storage cost of (i) eggs, butter, and cheese between 1924 and 1938, and (ii) wheat and oats between 1924 and 1932.

If anything, the decomposition results suggest that the storage cost component plays an important role.

Finally, our study contributes to the broader literature on the information content of the basis. This literature is directly motivated by the theory of storage. Fama and French (1987) document the predictive power of the futures–spot price spread for the changes in the spot price of several commodities. Brooks et al. (2013) extend this result to the oil market.[6] We complement this literature by showing that, consistent with the theory of storage, the *basis* predicts the spot return of the Gulf Coast sour crude oil. We go a step further and decompose the *basis* into the *scc* and *acyc*. This decomposition enables us to investigate the conduit through which the predictive power of the *basis* for spot returns arises. Our analysis reveals that the predictive power of the *basis* stems mostly from the *scc* rather than the *acyc*. To our knowledge, we are the first to document this result.

The remainder of this paper proceeds as follows. Section II presents the data and methodology. Section III discusses our main results. Section IV presents the additional robustness checks. Finally, Section V concludes.

II Data and Methodology

This section begins with an overview of our dataset. Next, the discussion focuses on the methodology.

[6]There is a related literature that analyzes the predictability of commodity futures risk premia with a particular focus on the cross-section. See Fuertes et al. (2010), Szymanowska et al. (2014), Koijen et al. (2018), and Boons and Prado (2019).

A Data

Storage Futures Contracts We analyze the LOOP SFC which started trading in March 2015. The SFC is a physically-settled futures contract that gives its holder the legal right to store 1,000 barrels of LOOP Gulf Coast sour crude oil at the Clovelly Hub during the delivery month.[7,8] It has a monthly delivery cycle spanning the next 15 consecutive months. For each delivery month, LOOP can offer a monthly storage capacity as high as 7.2 million barrels. Typically, the SFC is traded by commercial players such as oil refiners and oil producers as well as non-commercial players, including arbitrageurs engaging in cash-and-carry arbitrage as described in Ederington et al. (2021).

On the first Tuesday of every month, LOOP auctions new storage capacity either as SFC or physical forward agreement (PFA) on the proprietary Matrix Auction Platform.[9] Post-auction, the SFC trades on the Chicago Mercantile Exchange (CME) until the third business day before the twenty-fifth day of the month prior to the delivery month.[10] We focus on the SFC rather than the PFA for several reasons. To begin with, the secondary market of the PFA is over-the-counter, making it difficult to obtain the time-series of

[7] The LOOP crude oil storage facilities in Clovelly are the largest privately-owned crude oil storage facilities in the U.S. The LOOP facilities include 8 underground salt caverns that boast a total capacity of approximately 60 million barrels.

[8] It is important to stress that the salt cavern is the cheapest, and thus the most preferred, method of storage. One implication of this is that the associated storage cost is likely to be very responsive to market conditions. There are more expensive storage methods such as the onshore tanks, the railcars, as well as the offshore oil tankers. The Wall Street Journal reports that storing oil in an underground salt cavern *"can cost 25 cents a barrel each month, while storing crude on railcars costs about 50 cents a barrel and floating storage can cost 75 cents or more"*. We refer the interested reader to the following article: https://www.wsj.com/articles/the-new-oil-storage-space-railcars-1456655405.

[9] The Physical Forward Agreement (PFA) is a bilateral contract between LOOP and the holder of the PFA, which enables the latter to store 1,000 barrels of oil at Clovelly Hub.

[10] If the twenty-fifth calendar day of the month is not a business day, then trading ends on the third business day prior to the last business day preceding the twenty-fifth calendar day. For further details about the SFC, we refer the interested reader to the following page: https://www.cmegroup.com/trading/energy/crude-oil/loop-crude-oil-storage_contract_specifications.html.

prices needed for our main analyses. Second, the PFAs awarded in the auction need to be paid in full on the second business day following the award. This significant cash outlay means that only deep-pocketed investors tend to trade in this market. One upshot of this is that the auctions of the SFC are likely more competitive than those of the PFA.[11]

The LOOP SFC is a major innovation for the market of crude oil storage for at least three reasons. First, it makes the pricing of storage transparent to all market participants. Second, the monthly expiration schedule offers great flexibility to the end-users who were traditionally locked into long-term lease contracts. Third, because the SFC enables market participants to better manage their storage risk, it complements well the existing GCOFC, which is also physically settled at the Clovelly Hub. Given the preceding discussion, it is not surprising to see a growing number of storage venues launching similar products. For instance, in March 2019, the Intercontinental Exchange (ICE) launched its own ICE Permian WTI crude oil storage futures, which gives its holder the right to store 1,000 barrels of Permian WTI crude oil at Magellan's terminal in East Houston, Texas (MEH).[12] More recently, Matrix Global has also started running an auction for storage at the Oiltanking MOG'S Saldanha Bay facility in South Africa.[13]

We obtain the daily time-series of the settlement prices of the SFC for the period starting in March 2015 and ending in December 2019. Our dataset comes from Refinitiv Tick

[11] Indeed, conversations with market participants confirm this conjecture. For further details about the auction mechanism, we refer the reader to the following webpage: https://matrix.global/auctions/loop/.

[12] Magellan segregates storage for 500,000 barrels of oil for delivery over the first six months, and 1 million barrels of oil for the other maturities. For further details, please see https://www.theice.com/crude-oil/futures/permian-wti/qa.

[13] This facility consists of nine 1.121 million-barrel in-ground crude-oil storage tanks. https://matrix.global/auctions/otms/.

History and includes the settlement price, the settlement date, as well as the expiration date of each futures contract.[14] We process the dataset as follows. First, we discard the observations linked to the year 2015 since it corresponds to the first few months of trading of the SFC, when trading interest is likely low.[15] Second, we construct our monthly time-series of storage prices by sampling observations only on the last trading day of each futures contract.[16] In so doing, we essentially construct the monthly time-series of the cost of storing oil over the next month.

Gulf Coast Sour Crude Oil Futures We supplement our dataset of the SFC with the time-series of the GCOFC. This physically-settled crude oil futures contract has a monthly delivery cycle for the next 36 months. The underlying crude oil is a blend of Mars, Poseidon, and Segregation 17 crude oil streams. As is standard with crude oil futures contracts, the GCOFC is priced as a spread relative to the well-known West Texas Intermediate (WTI) crude oil futures.

Our dataset for the GCOFC comes from Refinitiv Tick History as well. We sample the observations of the GCOFC exactly on the same day as the SFC. Throughout this paper, we interpret the prices of the prompt and second nearby futures contracts as the spot and 1-month futures prices, respectively.

[14] This database has been used by several studies, including Rösch et al. (2017) and Hollstein et al. (2020).

[15] In a robustness check, we repeat our analysis after including the data related to the year 2015. Taking this step does not materially affect our results. We do not tabulate these results for brevity.

[16] Some studies roll the futures contracts at the end of the second month prior to delivery, e.g., Szymanowska et al. (2014), and sample observations at the end of the month. We do not follow this approach because our aim is to test the theory of storage. To do so, one needs an estimate of the spot price. Since, by no-arbitrage, a forward contract price must be equal to the spot price at maturity, we sample all observations on the last trading day of the expiring contract. This sampling scheme is consistent with the work of Fama and French (1987) who use the price of the maturing contract to proxy for the spot price.

It is important to stress that the GCOFC and the SFC are well-suited for our analysis since they are both physically-settled at the Clovelly Hub. To better understand this, suppose that a trader currently owns 1,000 barrels of oil that she wishes to carry forward and deliver in 1-month time to cover her short position in the expiring GCOFC. Since she must deliver 1,000 barrels of oil at the Clovelly Hub, we assume that the trader needs to pay for the 1-month storage at that location.[17] A convenient way to achieve this goal is for the trader to purchase the SFC to store oil until the delivery date. This example clearly shows that, by jointly using information about the GCOFC and the SFC, we can better analyze the cost-of-carry relationship.

B Methodology

Overview One of our goals is to explore the implications of the storage cost for the pricing of the crude oil futures contract. Our starting point is the cost-of-carry formula. This theoretical relationship posits that the futures price of a commodity to be delivered at a future date is equivalent to (i) purchasing the commodity today and (ii) incurring the carrying charges necessary to hold the inventory position until the delivery date.[18]

[17] We view this assumption as perfectly reasonable. Obviously, there are other methods of storage, e.g., onboard tankers, which are generally more expensive. Indeed, industry experts point out that the cost of storing oil aboard a very large crude carrier is generally 3 to 4 times higher than that of underground storage. It is important to note that, if one were to use these alternatives, the magnitude of the storage cost channel would be higher than our results suggest.

[18] Strictly speaking, this insight applies to the forward contract and not the futures contract. As is common in the literature, we simply assume that the forward price is the same as the futures price (Casassus and Collin-Dufresne, 2005).

We can formally write:

$$F_{t,t+1} = S_t + \underbrace{E_t\left(Storage_{t,t\to t+1} + X_{t,t\to t+1} - CY_{t,t\to t+1}\right)}_{\text{Carrying Costs}} \tag{1}$$

where $F_{t,t+1}$ is the price at time t of the futures contract for delivery at $t+1$. S_t is the spot commodity price at time t. $E_t(\cdot)$ denotes the expectation at time t. $Storage_{t,t\to t+1}$ is the cost, at time t, of storing the commodity over the period starting at t and ending at $t+1$. $X_{t,t\to t+1}$ denotes the cost, at time t, of all the other expenses for the period starting at t and ending at $t+1$. These other expenses include the administrative/handling fees, the insurance cost, the interest rate expense, and the pumping fees to name but a few. $CY_{t,t\to t+1}$ represents the convenience yield, at time t, related to the period starting at t and ending at $t+1$. As discussed in Casassus and Collin-Dufresne (2005), the convenience yield can be interpreted as a dividend flow that accrues to the inventory holder. Similar to Szymanowska et al. (2014), all carrying costs are (i) expressed in U.S. Dollars ($) and (ii) paid at the delivery date i.e., at time $t+1$.

By recording the price of the expiring SFC on its last trading day, we essentially observe the price, payable today, of storing oil for the next period. However, in the cost-of-carry relationship, *Storage* is paid at the end of the period of storage, rather than the beginning. Consequently, we need to compound the price of the expiring SFC using the corresponding interest rate. Thus, we can rewrite Equation (1) as:

$$F_{t,t+1} = S_t + \underbrace{SFC_{t,t\to t+1}(1 + r_{t,t\to t+1})^{1/12} + E_t\left(X_{t,t\to t+1} - CY_{t,t\to t+1}\right)}_{\text{Carrying Costs}} \tag{2}$$

where $SFC_{t,t\rightarrow t+1}$ is the price, at time t, of the storage futures contract that allows the holder to store crude oil over the period starting at t and ending at $t+1$. $r_{t,t\rightarrow t+1}$ denotes the annualized interest rate, at time t, for the period starting at t and ending at $t+1$. As is standard in the literature on derivatives, e.g., Avino et al. (2020), we use the term-structure of the London Interbank Offered Rate (LIBOR) to proxy for the interest rate.[19] We obtain this dataset from Bloomberg.

Re-arranging Equation (2), we compute the basis as in Boons and Prado (2019):

$$\underbrace{\frac{F_{t,t+1}-S_t}{S_t}}_{basis_t} = \underbrace{\frac{SFC_{t,t\rightarrow t+1}(1+r_{t,t\rightarrow t+1})^{1/12}}{S_t}}_{scc_t} - \underbrace{E_t\left(\frac{CY_{t,t\rightarrow t+1}-X_{t,t\rightarrow t+1}}{S_t}\right)}_{acyc_t} \quad (3)$$

$$basis_t = scc_t - acyc_t \quad (4)$$

where $basis_t$ stands for the basis at time t. scc_t and $acyc_t$ indicate the storage cost and the adjusted convenience yield channels at time t, respectively.

Implications Equation (3) clearly shows that the *basis* captures the difference between the *scc* and the *acyc*. This expression also makes a number of interesting predictions regarding the time-series properties of the *basis*.

First, the average of the *basis* can be decomposed into the following two components:

$$E(basis) = E(scc) - E(acyc) \quad (5)$$

[19] As a robustness check, we use the term-structure of the Treasury rates as a proxy for the interest rate. As Section IV.C shows, the results are similar. This finding is not surprising given (i) our interest in the short-term futures contract and (ii) the low interest rate regime that prevails over our sample period.

Assuming that the average of the basis is not equal to 0, we can re-arrange Equation (5) as follows:

$$100\% = \frac{E(scc)}{E(basis)} - \frac{E(acyc)}{E(basis)} \tag{6}$$

Equation (6) shows that we can decompose the average of *basis* into a component linked to the average (i) *scc* and (ii) *acyc*. This insight is important because the literature is unclear about the impact of the storage costs on the level of the basis. On the one hand, Ederington et al. (2021) argue that the storage costs are too small. On the other hand, Gu et al. (2020) leave open the possibility that the storage costs might be important for the level of the basis. Because these studies do not have data about the pricing of commodity storage, they cannot provide quantitative estimates of the impact of the cost of storage on the level of the basis.

Second, Equation (4) sheds light on the impact of the storage cost channel on the variations in the basis. To see this, notice that:

$$\begin{aligned} Var(basis) &= Var(scc - acyc) \\ 100\% &= \underbrace{\frac{Var(scc) - 2 \times Cov(scc, acyc)}{Var(basis)}}_{\text{Variance Contribution}_{scc}} + \underbrace{\frac{Var(acyc)}{Var(basis)}}_{\text{Variance Contribution}_{acyc}} \end{aligned} \tag{7}$$

Equation (7) decomposes the variance of the *basis* into 2 components. The first component comprises (i) the variance of the *scc* and (ii) the covariance of the *scc* and *acyc*. The second component is solely driven by the time-variations in the *acyc*.

One may wonder why we assign the covariance term to the contribution of the *scc*. To understand our motivation, it is useful to recall that the extant literature argues that the cost of storage displays very little time-variations (Gu et al., 2020). Under the null hypothesis of a constant storage cost, the first component to the right of the equality sign of Equation (7) will be equal to zero, leaving all the variations in the *basis* to the *acyc*. If this hypothesis is rejected by the data, then Equation (7) provides a framework to quantify the contribution of the fluctuations in the cost of storage to the variability of the basis.

Implementation In order to operationalize the decompositions in Equations (6) and (7), we need to clarify the computation of the key variables in Equation (4). Using the time-series of oil spot and futures prices, we compute the *basis* as:

$$basis_t = \frac{F_{t,t+1} - S_t}{S_t} \tag{8}$$

Next, we use the monthly time-series of (i) the price of the expiring SFC, (ii) the interest rate, and (iii) the crude oil spot price to compute the *scc* as:

$$scc_t = \frac{SFC_{t,t\to t+1}(1 + r_{t,t\to t+1})^{1/12}}{S_t} \tag{9}$$

Finally, we estimate the *acyc*. This estimation is complicated by the fact that the convenience yield and other carrying costs are not directly observable. To by-pass this problem, we adopt a residual-based modelling framework.[20] That is, we re-arrange Equation (4) to express the *acyc* channel as the difference between two observable quantities, namely the *scc* and the *basis*:

$$acyc_t \quad = \quad scc_t - basis_t \tag{10}$$

The advantage of this approach is that it directly imposes the economic restriction implied by the identity in Equation (4). One might be tempted to posit a reduced-form model for the *acyc*. We refrain from pursuing this approach for several reasons. First, this channel is not directly observable. This can be problematic because, in order to estimate the parameters of the model, the econometrician will need to make an assumption regarding the correct proxy. In turn, this assumption makes the estimation results sensitive to the quality of the proxy. Second, by estimating a time-series model, one cannot guarantee that the identity presented in Equation (4) always holds.

III Main Results

This section presents our main results. To begin with, we discuss the time-series properties of the key variables. Next, we empirically decompose the *basis* into the *scc* and the *acyc*. Finally, we explore the information content of the basis and investigate the source(s) of its predictive power.

[20] The residual-based modelling approach is pervasive in the literature. See Campbell and Shiller (1988) for an early example of this approach.

A Overview

Before turning to our main results, it is instructive to inspect the time-series behaviour of our key variables. Figure 1 reveals important time-variations in the *basis* with the shaded bars indicating the periods of contango. We define the contango (backwardation) periods as periods associated with a positive (negative) *basis*. During contango periods, inventory holders prefer to store the commodity and carry the position forward. We can see that the GCOFC is in contango about half of the time.

The top panel of Figure 2 shows that the *scc* displays considerable time-series variations, as evidenced by values that range from 0.02% to close to 4.85% per month.[21] In order to get more insights into these time-variations, we compare the periods of contango and backwardation. Since contango months correspond to periods when inventory owners have a stronger incentive to store oil until market conditions improve, we expect the *scc* to take higher values during these times. The top panel of Figure 2 presents evidence consistent with this conjecture.

The bottom panel of Figure 2 depicts the dynamics of the *acyc*. We can see that the *acyc* sometimes take negative values over our sample period. At first glance, this result seems puzzling as one would expect the *acyc* to be bounded from below by zero. However, recall that our *acyc* is the spread between the convenience yield and the other costs of carrying the inventories forward (see Equation (3)). Clearly, if the convenience yield is

[21] The reader may notice that the *scc* takes large values in the first quarter of 2016. To understand this result, it is useful to recall that the *scc* is positively related to the SFC and negatively related to the spot price of oil. During the first quarter of 2016, the price of oil is particularly low, thus boosting the *scc*. Furthermore, owners of the physical asset react to low prices by storing the commodity, thereby exerting an upward pressure on the price of the SFC, which in turn results in a higher *scc*.

equal to zero, these other carrying charges can make the *acyc* turn negative.

Table 1 summarizes the main statistics for the *basis*, the *scc*, and the *acyc*. Several results are worth pointing out. On average, the *basis* is positive as evidenced by its sample mean of 0.18 % per month. It also displays substantial time-variations, fluctuating between −4.93% and 6.16% over our sample period. The average *scc* is equal to 0.50% per month. From an economic standpoint, this result sheds some light on the cost incurred by traders participating in the physical market.[22]

We can also see that the *scc* displays a volatility of 0.89 % per month, which is higher than its mean. This volatility estimate challenges the conventional assumption that the storage cost is constant in the time-series dimension (Gu et al., 2020). Furthermore, we can see that the *scc* displays a skewness of 3.28 and a kurtosis of 14.64, suggesting that it is subject to extreme positive movements. Because of (i) the small sample size and (ii) the high values of the higher-order moments, we need to be careful when carrying out our statistical inference.[23]

Finally, we note that the *acyc* is positive on average, suggesting that the convenience yield generally dominates all the other carrying costs, e.g., insurance and interest expenses.[24] Analyzing the variance of this channel, we can see that it is more volatile

[22]One may quibble that our interpretation of the *scc* as an estimate of the proportional cost of storage is not quite accurate since we compound the SFC price using the interest rate (see Equation (3)). This compounding effect makes the *scc* an upward biased estimate of the proportional storage cost. However, it is worth pointing out that the compounding effect is negligible given (i) our focus on the month-ahead and (ii) the low interest rate environment that prevails over our sample period.

[23]To alleviate this concern, we compute the standard errors using the dependent wild bootstrap of Shao (2010). Section IV.B presents the results of the bootstrap experiment and confirms that our key results hold. As an additional robustness check, we winsorize the observations at the 95% level. This untabulated analysis yields qualitatively similar results.

[24]Given our focus on the 1-month horizon and the low interest rates that prevail during our sample period, we anticipate that the contribution of interest rate expenses to both the mean and the variance of the *acyc* is likely small.

(1.38 %) than the *scc*. Also, its higher-order moments are closer, compared with those of the *scc*, to those of a normal distribution.

B Decomposing the Basis

Unconditional Analysis We decompose the mean and variance of the *basis* as per Equations (6) and (7). Panel A of Table 2 shows that the *scc* contributes 281.05% to the mean of the *basis* while the *acyc* contributes 181.05%.[25] Therefore, the *scc* accounts for a sizeable share of the mean of the *basis*. To the extent that our results can be generalized to other markets, this finding gives credence to the hypothesis that cross-sectional sorts of commodities based on their carry signal, e.g., Fuertes et al. (2010), are potentially driven by cross-sectional differences in storage costs (Gu et al., 2020).[26]

Turning to the variance decomposition, we can see that the *scc* contributes close to half (45.35%) of the variation in the *basis*. This finding challenges the assumption that storage costs display very little variations in the time-series dimension (Gu et al., 2020; Ederington et al., 2021). Since the contribution of the *scc* to the variance of the *basis* depends on the variance of the *scc* and the covariance of the *scc* with the *acyc* (see Equation (7)), one may wonder about the relative importance of these two terms. To shed light on this, it is useful to recall that Table 1 reports the standard deviation estimates of the *scc* (0.89%)

[25]Note that these contributions are greater than 100 % simply because the average *basis* (0.18%) is lower than the averages of (i) the *scc* (0.50%) and (ii) the *acyc* (0.32%).

[26]Fama and French (1987) present some evidence that may be consistent with this effect. Table 2 of their paper reports the estimates of storage costs for a broad range of commodity markets. The storage cost figures relate to June 1984 and are expressed as a percentage of the spot price. We download the commodity futures price and 3-month Treasury rate data for June 1984 from Bloomberg. We then compute the *basis* and *scc* based on Equations (3) and (9), respectively. For each commodity market, we compute the contribution of the *scc* to the *basis* as the ratio of the *scc* over the *basis* of the corresponding period. Our untabulated analysis points to large cross-sectional variations in the contribution of the *scc* to the level of the *basis* with an estimate as high as 222% in the wheat market.

and the *basis* (1.87%). Equipped with these statistics, we can calculate the contribution of the variance of the *scc* to the total variance of the *basis*. Straightforward calculations indicate that the variance of the *scc* alone accounts for 22.65% of the variance of the *basis*. The upshot of this is that the *scc* and *acyc* are negatively correlated and their covariance captures 22.70% of the variance of the *basis*.

Conditional Analysis It is interesting to analyze the results over different market conditions. A priori, one would expect the *scc* (*acyc*) to become more (less) important when inventory owners have an incentive to carry their position forward. We thus repeat our decomposition for different market states. In order to identify these market states, we rely on the periods of backwardation and contango.

Panel B of Table 2 shows that the *scc* accounts for a small share of the mean and variance of the basis during backwardation periods.[27] This result is to be expected. After all, inventory holders have very little incentive to store oil during backwardated markets. Panel C of the same table shows that the *scc* becomes more dominant during contango periods, when the incentive to store is stronger. It shows that the *scc* is the main contributor to the (i) mean (62.36%) and (ii) variance (76.34%) of the basis.[28]

[27] The negative sign of the contribution of the *scc* to the mean of *basis* arises from the fact that the *scc* is by definition positive, while the mean *basis* is negative during the backwardation period.

[28] One may wonder about the relatively high contribution (−37.64 %) of the *acyc* to the mean *basis* during the contango periods. A priori this channel should be equal to 0 during contango periods since the convenience yield is expected to be negligible during these times. While this is true, it is important to remember that the *acyc* is the difference between the convenience yield and other expenses, including insurance and handling costs. It is plausible that the other carrying costs are positively correlated with the storage costs, which increase during periods of high inventories, thus making the *acyc* take negative values.

C The Information Content of *scc* for Future Inventories

Overview The *scc* is based on the SFC which is informative about the price of storing oil during the delivery month. To the extent that this futures contract is informative about future inventory levels, the *scc* should predict the future stock levels.

To test this prediction, we download inventory data from the website of the Energy Information Agency (EIA). The EIA is an authoritative source of inventory data that has been used in several studies, e.g., Baker (2021) and Mukherjee et al. (2021). Specifically, we obtain the commercial crude oil stock level (expressed in 1,000 barrels), excluding the lease stock, pertaining to the Petroleum Administration for Defense Districts 3 (PADD3).[29] We focus on the PADD3 region because it covers the Gulf Coast area, which includes the storage facilities of LOOP. More specifically, the inventory dataset includes the domestic and customs-cleared foreign inventories currently at, or in transit to, refineries and bulk terminals, and inventories in pipelines.[30] Although the stock levels are as of the Friday of each week, the inventory report is only published on the following Wednesday. This implies that there is a publication delay of 5 days (Armstrong et al., 2021; Prokopczuk et al., 2021). Throughout this paper, our main timing convention is that the

[29] The U.S. is divided into 5 PADD regions. For more details on the geographical split of regions into different PADD areas, we refer the interested reader to the following page: `https://www.eia.gov/todayinenergy/detail.php?id=4890`. Mukherjee et al. (2021) show that the PADD3 region is important in that it accounts for 50.21% of the U.S. inventories as of the end of December 2016. Our weekly inventory dataset of the PADD3 region is available from the following webpage: `https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=WCESTP31&f=W`.

[30] As previously discussed, we take the perspective of an investor who stores oil at the Clovelly Hub. Thus, we do not consider the inventories onboard tankers. Viewed in this light, the EIA inventory data is well-suited for our purposes as it does not include information about inventory levels onboard tankers or at international locations that could easily be shipped to the PADD3 area. See Gorton et al. (2013) for a similar point about the inventory data of WTI crude oil which belongs to the PADD2 region.

inventory data is associated with its publication date rather than its measurement date.[31] Equipped with this dataset, we compute the inventory growth rate as:[32]

$$\%\Delta Inv_{t+1} \quad = \quad \frac{Inv_{t+1} - Inv_t}{Inv_t} \tag{11}$$

where $\%\Delta Inv_{t+1}$ is the 1-period growth rate in the inventory data at time $t+1$. Inv_{t+1} and Inv_t denote the most recent inventory data published by times $t+1$ and t, respectively.

Since we are modelling the growth rate in inventories, we also define the growth in the scc as:

$$\%\Delta scc_t = \frac{scc_t - scc_{t-1}}{scc_{t-1}} \tag{12}$$

where $\%\Delta scc_t$ is the growth rate in the storage cost channel observed at time t.

We run a regression of the 1-period inventory growth rate on a constant and the lagged

[31] As a robustness check, we consider the timing convention of Ederington et al. (2021) who assume that the inventory data is available directly on the Friday. As Section IV.A shows, this alternative choice leads to similar results.

[32] Alternative approaches exist. For instance, Alquist et al. (2014) model the logarithmic growth of the inventories. Another possibility is to predict the change, rather than the growth rate, of inventories (Ederington et al., 2021). In robustness checks, we consider both of these approaches and find that they lead to similar results. See Tables A1 and A2 of the Online Appendix.

$\%\Delta scc$:[33]

$$\%\Delta Inv_{t+1} = \alpha + \beta \times \%\Delta scc_t + \epsilon_{t+1} \quad (13)$$

where α and β are the intercept and slope parameters, respectively. If $\%\Delta scc$ contains information about the future inventory growth rate, we expect a positive and significant slope parameter estimate. Accordingly, we carry out a 1-sided significance test. Throughout the paper, we use the 5% significance level. We always report the Newey and West (1987) standard errors computed with 2 lags in parentheses.

Table 3 summarizes the regression results. The first set of results from the left reveals that $\%\Delta scc$ enters the regression with a positive and statistically significant slope estimate (0.020, t-stat=2.802). The explanatory power of this regression (Adj R^2=7.07%) is sizeable too, confirming that the $\%\Delta scc$ contains information about the future inventory growth rate.

This finding is highly relevant for studies that connect the basis to inventories. Alquist et al. (2014) document that the principal components extracted from the term-structure of the slopes of the WTI crude oil futures predict the future inventory growth rate.[34]

[33] There is very little evidence in the literature to suggest that crude oil inventory data follow a seasonal pattern at the monthly sampling frequency. Gorton et al. (2013) regress the monthly normalized inventory data on 12 monthly dummy variables. The authors find that, among all energy commodities analyzed, the model displays the lowest explanatory power for the WTI crude oil market (see their Table II). Symeonidis et al. (2012) confirm these findings, writing that there is no indication of seasonality in either the basis or the inventory data linked to the WTI crude oil. While these papers focus on the WTI oil market, there is no guarantee that the findings apply to our market. We therefore perform a similar analysis on the time-series of the monthly PADD3 inventory data. We do not find any indication of seasonality.

[34] There are several differences between our work and that of Alquist et al. (2014). To begin with, the authors focus on the term-structure of WTI crude oil futures. Furthermore, they regress the inventory growth rate on the levels of the principal components. Our analysis focuses on the Gulf Coast sour crude oil and involves regressions of growth rates on growth rates.

However, in their framework, it is not clear which of the *scc* and the *acyc* accounts for most of the documented predictive power. In order to further shed light on this, we augment our baseline model with the lagged growth rate of the *acyc* ($\%\Delta acyc$). Since the *acyc* can be interpreted as the benefit, net of other carrying costs, accruing to the inventory holder, we anticipate a negative relationship between $\%\Delta acyc$ and the future inventory growth rate.

In other specifications, we include the lagged inventory growth rate and control for several variables that may also have an impact on the growth rate of inventories.[35] We build on Ederington et al. (2021) and select three additional variables.[36] The first control variable is the growth rate of the total imports into PADD3. Intuitively, if there is an unexpected increase in the quantity of imported oil, we should see a higher inventory growth rate.[37] The second variable relates to the growth rate of the quantity of crude oil used as input by refineries.[38] If refineries unexpectedly use more crude oil as input into their production process, we expect to see a lower growth rate of inventories. The third

[35]We also consider the logarithmic growth of these variables. The only exception to this is the *basis* since it can take negative values. For that variable specifically, we retain the simple growth rate. Table A1 of the Online Appendix presents results that are similar to our baseline estimates.

[36]Similar to Ederington et al. (2021), we use the control variables observed at time $t+1$. Implicit in this analysis is the assumption that the econometrician is able to perfectly predict the next-period value of the control variables. Clearly, this is a strong assumption. In an untabulated regression, we include the control variables at time t, rather than $t+1$, and repeat our main analyses. We find that the main conclusion regarding the information content of the $\%\Delta scc$ is unchanged. These results are available on request.

[37]Ideally, we would want to study the imports net of exports. Although this dataset is available for the whole of the U.S., it is unfortunately not available at the PADD level. Hence, we simply restrict our attention to the imports, expressed in thousands barrels per day. The dataset is available at the following address: `https://www.eia.gov/dnav/pet/pet_move_wkly_dc_NUS-Z00_mbblpd_w.htm`.

[38]The quantity of crude oil used as refinery inputs is expressed in thousands barrels per day. For further information about this dataset, we refer the interested reader to the following page: `https://www.eia.gov/dnav/pet/pet_pnp_wiup_dcu_r30_w.htm`.

variable relates to the growth rate of the U.S. production of oil.[39] Holding everything else constant, an unexpected increase in the U.S. oil production should result in higher inventory growth. The third column of Table 3 shows that the slope estimate associated with $\%\Delta scc$ (0.020) and its significance (t-stat=2.157) are very similar to those of our benchmark estimation (0.020, t-stat=2.802). Furthermore, we can see that, consistent with our intuition, the loading on the $\%\Delta acyc$ is negative and statistically significant (-0.001, t-stat=-3.951).

We also interact the $\%\Delta scc$ with the backwardation and contango dummy variables. We can see from Table 3 that the information content of the channel linked to the $\%\Delta scc$ is stronger during periods when the underlying oil market is in contango. While the slope observed during periods of backwardation is not significant (0.009, t-stat=1.877), we observe a statistically significant estimate (0.042, t-stat=4.379) during periods of contango. Similar to our benchmark estimates, controlling for other variables does not change our main message: the information content of the storage cost is stronger during periods when the incentive to store oil is high.

D The Information Content of the SFC for future Spot Prices

We analyze the information content of the *basis* for future spot returns:

$$R_{t+1} = \alpha + \beta \times basis_t + \epsilon_{t+1} \tag{14}$$

[39] The dataset shows the U.S. field production of crude oil in thousands of barrels per day. The complete information can be retrieved from https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=WCRFPUS2&f=W.

where R_{t+1} denotes the simple return on oil at time $t+1$, computed as $R_{t+1} = \frac{S_{t+1}-S_t}{S_t}$.[40] If the *basis* is informative about future spot returns, we should observe a statistically significant slope estimate. Table 4 shows that the loading on the *basis* is positive and statistically significant (2.241, t-stat=2.382).[41]

Since *basis* can be decomposed into the *scc* and *acyc* (see Equation (4)), it is interesting to investigate the contribution of each of these channels to the predictability results. We start by separately estimating univariate regressions of spot returns on a constant and each of these channels:

$$R_{t+1} = \alpha + \beta \times scc_t + \epsilon_{t+1} \quad (15)$$

$$R_{t+1} = \delta + \gamma \times acyc_t + \epsilon_{t+1} \quad (16)$$

Intuitively, we expect the loading on the *scc* and *acyc* to be positive and negative, respectively. The reason is that the *basis* depends positively on the *scc* and negatively on the *acyc* (see Equation (3)). The univariate regression results of Table 4 reveal that the *scc* predicts future spot returns with a positive and statistically significant slope estimate (4.475, t-stat=3.014). The explanatory power of this regression (Adj R^2=13.87%) is very similar to that of the regression of the spot returns on a constant and the lagged *basis* (Adj

[40] One may wonder whether our empirical results are robust to the use of logarithmic returns. Table A3 of the Online Appendix presents the results based on the log returns. Generally, these results are consistent with those based on simple returns.

[41] The reader may wonder whether the *basis* is informative about the futures risk premium too. Building on Fama and French (1987), it is straightforward to show that the *basis* may be informative about the future spot return and the futures risk premium. The slope parameter of Equation (14) is informative about the share of variation in the *basis* associated with the future spot returns. Table 4 reveals that we cannot reject the null hypothesis that $\beta = 1$. We thus conclude that the *basis* is informative about future spot returns rather than the futures risk premium. This result extends the finding of Brooks et al. (2013), who study the WTI crude oil market, to the Gulf Coast sour crude oil market.

R^2=15.04%). The economic intuition is simple: market participants are likely to pay more for storage when they expect higher future spot returns. Turning to the predictive power of the *acyc*, we find that the slope estimate is negative but not statistically significant (−2.196, t-stat=−1.654).[42]

Next, we estimate an encompassing regression that includes both the *scc* and *acyc* as forecasting variables:

$$R_{t+1} = \alpha + \beta \times scc_t + \gamma \times acyc_t + \epsilon_{t+1} \tag{17}$$

Table 4 shows that the loading on the *scc* is positive and highly significant (3.800, t-stat=3.334). This slope estimate is comparable to that observed in the univariate regression of the spot return on a constant and the lagged *scc* (4.475, t-stat=3.014). Turning to the *acyc*, we find that its slope parameter estimate is insignificant (t-stat=−1.292). Furthermore, the explanatory power of this encompassing regression (Adj R^2=15.40%) is very similar to that of the model that includes the *scc* as the only forecasting variable (Adj R^2=13.87%). Taken together, the evidence suggests that the *scc*, rather than the *acyc*, is the main conduit through which the predictive power of the *basis* arises. We are the first to document this result. To the extent that this result extends to other commodity markets, it poses a challenge to the conventional wisdom in the literature that the predictive power of the commodity futures *basis* is driven by the component linked to the convenience yield.

[42] It is worth mentioning that, if one were interested in a 1-sided alternative hypothesis test based on the hypothesized negative relationship, the critical value would be −1.679. The upshot of this is that the *acyc* does not predict future spot returns.

Although consistent with the insights of the theory of storage, one may criticize the prior analysis on the ground that it does not control for other potentially important return forecasting variables. To mitigate this concern, we control for a number of variables used in Gu et al. (2020).[43] Specifically, we use the relative basis (*relbasis*), momentum (*mom*), and the basis momentum (*basmom*) defined as:

$$relbasis_t = \frac{S_t - F_{t,t+1}}{F_{t,t+1}} - \frac{F_{t,t+1} - F_{t,t+2}}{F_{t,t+2}} \tag{18}$$

$$mom_t = \Pi_{i=t-11}^{t} \left(\frac{S_i}{F_{i-1,i}} \right) - 1 \tag{19}$$

$$basmom_t = \Pi_{i=t-11}^{t} \left(\frac{S_i}{F_{i-1,i}} \right) - \Pi_{i=t-11}^{t} \left(\frac{F_{i,i+1}}{F_{i-1,i+2}} \right) \tag{20}$$

where $F_{t,t+2}$ is the time-t price of the futures contract expiring at $t+2$. The last three columns of Table 4 show that including these variables does not significantly affect the information content of the *scc*.

IV What About ...

Having documented our main results, we carry out a number of additional tests. First, we consider a different timing convention for the inventory data. Second, we conduct a bootstrap experiment to ensure that our statistical evidence is robust. Third, we evaluate the impact of the choice of the interest rate proxy on our main results. Fourth, we use an alternative definition of market conditions that is based on the spare capacity level.

[43] In interpreting our evidence, one should be careful about the fact that most of the literature focuses on the predictability of the futures risk premium rather than the spot return. As a result, the forecasting variables in Gu et al. (2020) are not necessarily well-suited for a study of spot return predictability.

Fifth, we investigate the extent to which the convexity variable of Gu et al. (2020) may be useful to purge the slope between the futures and spot prices from the effect of the storage cost. Sixth, we explore whether the *scc* may be informative about the WTI spot oil return. Seventh, we examine the predictive power of the *scc* for the stock returns of companies active in the mid-stream segment of the energy value chain.

A Timing Convention?

Our analysis of the predictability of the inventory growth rate accounts for the publication delay inherent to inventory news. However, Rousse and Sévi (2019) study the dynamics of the WTI futures prices and present evidence which suggests the possibility of information leakage prior to the official inventory announcement. This result may explain why Ederington et al. (2021) use a timing convention based on the measurement date, i.e., the Friday, rather than the release date.

As a robustness check, we assume that the inventory data is known to market participants on Friday as soon as the measurement is completed and repeat our analysis. Table A4 of the Online Appendix shows that the results are similar to our benchmark findings.

B The Statistical Inference?

Given that the SFC is a relatively new instrument, our sample period is necessarily short. Furthermore, the summary statistics of Table 1 reveal that the distribution of the *scc* is positively skewed and displays excess kurtosis. One may naturally wonder about the effect of the aforementioned facts on our statistical inferences. To address this concern, we implement the dependent wild bootstrap of Shao (2010). This approach preserves the linear dependence present in the original data and allows for heteroskedasticity.

To fix ideas, suppose that the regression model of interest is that of Equation (13). We draw a vector of residuals, of the same length as the time-series of our dependent variable, from a standard normal distribution. We model the dependence in the residuals through a Bartlett kernel using a bandwidth parameter equal to 2.[44] We then multiply each of the generated random series with the corresponding residual of the regression (see Equation (13)), thus obtaining the time-series of bootstrapped residuals. Next, we use the bootstrapped residuals, the estimated parameters, and the forecasting variables to generate the time-series of the bootstrapped dependent variable. We repeat the previous steps 9,999 times to obtain 9,999 pseudo-samples of the dependent variable of the forecasting regression. For each pseudo-sample, we then use the bootstrapped series to estimate the forecasting regression. As a result, we obtain 9,999 estimates of each parameter. For each parameter, we obtain the bootstrapped standard error by computing the standard deviation across the 9,999 estimates.

We use the bootstrapped standard error to compute the bootstrapped t-stat which we report in square brackets (see Tables 3 and 4). Inspecting these quantities, we can see that they are similar to our baseline Newey and West (1987) t-statistics. This result gives us more confidence in our analysis.

[44]Our interest in this kernel and bandwidth is mainly motivated by the desire to coincide with the Newey and West (1987) estimator computed using 2 lags.

C The Proxy of Interest Rates?

When computing the *scc* (see Equation (9)), we use the term-structure of the LIBOR rates as a proxy for the riskless rate. While this approach is standard in the literature, it is nonetheless important to assess the sensitivity of our results to this assumption. As an alternative, we use data on the term-structure of the U.S. Treasury interest rate, which we download from Bloomberg.

Tables A5–A8 of the Online Appendix summarize the results. Generally, these results are very similar to our baseline estimates and confirm our key findings. The *scc* makes a substantial contribution to both the mean and variance of the *basis*. This set of findings is more pronounced during the contango periods. Furthermore, we can see that the *scc* is the main conduit through which the predictive power of the *basis* emerges.

D Different Definitions of Market States?

So far, we have relied on the sign of the *basis* to identify the market states. Since the *basis* takes positive values when the futures price is higher than the contemporaneous spot price, one might worry that this will mechanically lead to a higher *scc* during these times. To address this concern, we now consider a proxy of market states based on the spare capacity level at the PADD3 district. The computation of this measure is as follows. First, we retrieve the time-series of the weekly PADD3 stocks, excluding the strategic petroleum reserves of crude oil. Second, we deduct from this dataset the quantity of oil in pipeline fills or in transit by water and rail. Third, we divide the resulting figure by the refinery, tank, and underground working storage capacity. The final estimate gives us a measure of the used storage capacity. The spare capacity is simply the difference between 1 and the used storage capacity.

Equipped with this dataset, we now identify 2 market states. The first market state relates to periods during which the spare storage capacity is below the median spare capacity computed using the full sample. We call this the low spare capacity state. The second market state, which we term the high spare capacity, relates to all other periods. Intuitively, we expect that the price of storage should be greater when there is very low spare capacity available. Table A9 of the Online Appendix decomposes the basis for each of these two regimes. It shows that the *scc* dominates both the mean (83.16%) and the variance (64.70%) of the *basis* during the low spare capacity regime.

We also repeat our analysis of inventory growth predictability after conditioning on the state of the market. As the last two columns of Table 3 show, our results are similar to the baseline estimates. The variables linked to the *scc* strongly predict future inventory growth rates. Most of this predictive power arises from periods when there is limited spare capacity in the market.

E The Convexity (Gu et al., 2020)?

In a recent paper, Gu et al. (2020) point out that storage costs may significantly contribute to the level of the basis. In turn, this effect implies that cross-sectional sorts on the basis may be affected by cross-sectional differences in the storage costs, complicating the economic interpretation of the findings of papers analyzing the carry trade in commodity markets (Fuertes et al., 2010). The authors assume that the storage cost is constant over time and propose to compute the convexity, which is defined as the difference between the first two bases:

$$convexity_t = \frac{F_{t,t+2} - F_{t,t+1}}{F_{t,t+1}} - \frac{F_{t,t+1} - S_t}{S_t} \tag{21}$$

where $convexity_t$ is the convexity measure at time t.

Straightforward computations reveal that:

$$\begin{aligned} convexity_t &= \underbrace{\frac{SFC_{t,t+1\to t+2}(1 + r_{t,t+1\to t+2})^{1/12}}{F_{t,t+1}} - \frac{SFC_{t,t\to t+1}(1 + r_{t,t\to t+1})^{1/12}}{S_t}}_{scc_t^{convexity}} \\ &\quad - \underbrace{\left(E_t\left(\frac{CY_{t,t+1\to t+2} - X_{t,t+1\to t+2}}{F_{t,t+1}}\right) - E_t\left(\frac{CY_{t,t\to t+1} - X_{t,t\to t+1}}{S_t}\right)\right)}_{acyc_t^{convexity}} \\ convexity_t &= scc_t^{convexity} - acyc_t^{convexity} \end{aligned} \tag{22}$$

where $SFC_{t,t+1\to t+2}$ is the price at time t of the storage futures contract that enables its owner to store the commodity from $t+1$ to $t+2$. $r_{t,t+1\to t+2}$ is the forward interest rate at time t pertaining to the period that starts at $t+1$ and ends at $t+2$. $scc_t^{convexity}$ and $acyc_t^{convexity}$ denote the values, at time t, of the storage cost and adjusted convenience yield components of *convexity*, respectively. We compute $scc_t^{convexity} = \frac{SFC_{t,t+1\to t+2}(1+r_{t,t+1\to t+2})^{1/12}}{F_{t,t+1}} - \frac{SFC_{t,t\to t+1}(1+r_{t,t\to t+1})^{1/12}}{S_t}$. Lastly, we obtain $acyc_t^{convexity}$ as the difference between $convexity_t$ and $scc_t^{convexity}$.

Equation (22) reveals that, if (i) the short-end of the term-structure of the SFC is flat, (ii) the short-end of the term-structure of interest rates is constant, and (iii) the *basis* is equal to zero, i.e., $F_{t,t+1} = S_t$, then the $scc^{convexity}$ will be equal to zero. One upshot of this is that *convexity* will effectively be free of the influence of the storage cost as predicted by Gu et al. (2020). We can check whether these assumptions are borne out by the data by combining our dataset of SFC and the insights of Equation (22) to gauge the contribution of $scc^{convexity}$ to the dynamics of *convexity*.

Table 5 decomposes the mean and variance of the *convexity*. We can see that the $scc^{convexity}$ dominates the mean (289.93%) of the *convexity* and makes a modest contribution (17.69%) to the variance of the *convexity*. Taken together, the results suggest that the *convexity* approach of Gu et al. (2020) does not completely eliminate the effect of the storage costs.

F The Predictability of WTI Spot Oil Returns?

So far, our analysis focuses on the predictability of the spot returns on the Gulf Coast crude oil. Our interest in this market is directly motivated by the fact that it is well-aligned with our dataset of SFC and GCOFC. Nonetheless, the WTI is the benchmark crude oil market and there are pipelines connecting the PADD3 area to the PADD2 area, which includes Cushing in Oklahoma, the main delivery point of the WTI crude oil. This insight raises the possibility that the *scc* may, to some extent, comove positively with the cost of storing oil in Cushing. That is, the storage cost in the Clovelly Hub may be a (noisy) proxy for the storage cost in Cushing. Under this assumption, our *scc* may be informative about the WTI spot oil returns.

A few caveats are worth discussing. First, the analysis implies that there are no significant frictions, e.g., transportation costs, hindering the flow of oil between the PADD3 and PADD2 regions. Clearly, if this assumption does not hold, then we will struggle to find evidence of predictability. Second, the storage cost at the Clovelly Hub is a noisy proxy for that of Cushing. This noise potentially introduces measurement errors in the analysis, leading to an attenuation bias. Put simply, the noise works against us and makes it difficult to find evidence of predictability.

Table 6 reveals that the *scc* is a strong predictor of WTI spot returns. In univariate regressions, it is associated with a significantly positive slope parameter (3.920, t-statistic=2.739). In the encompassing regression that includes the *scc*, the *acyc*, and the same control variables as in our analysis of the predictability of the sour crude oil spot returns, we find that its slope coefficient is still positive and highly significant (3.023, t-stat=2.675).[45] Remarkably, this analysis reveals that, notwithstanding the attenuation bias inherent to this analysis, the *scc* contains information about the benchmark WTI crude oil spot return.

G The Predictability of the Returns on Mid-Stream Companies?

We push our predictability analysis a step further. Specifically, if the SFC is truly informative about the pricing of storage, we should see that it predicts the stock returns of companies that specialize in the transportation and storage of crude oil. To shed

[45] One may object to this analysis on the grounds that our control variables are based on data linked to the GCOFC. To address this concern, we compute the control variables based on the WTI futures contract. Table A10 of the Online Appendix presents results that are similar to our baseline findings.

light on this, we focus on the ETFs that track the Alerian MLP infrastructure index.[46] This capitalization-weighted index comprises energy firms, headquartered in the U.S. and Canada, that generate most of their cash flows from mid-stream activities such as transportation and storage.[47] We obtain the time-series of prices of the (i) Alerian Energy Infrastructure (ENFR), (ii) VanEck Vectors (EINC), and (iii) Alerian MLP (AMLP) ETFs from Bloomberg. All three ETFs track the same Alerian MLP index.[48] Table 7 summarizes the results of the regression of the monthly returns on each ETF on a constant and the lagged *scc*. We can see that, consistent with our hypothesis, the *scc* exhibits strong predictive power for the future returns on each ETF.

V Conclusion

We use a novel dataset of SFC to conduct the first detailed analysis of the dynamics of the cost of oil storage. We find that the 1-month storage cost represents on average 0.50% of the spot price of oil and displays considerable time-variation. These fluctuations challenge the common assumption in the literature that the storage cost varies little over time. We develop and implement a novel decomposition of the *basis* into the *scc* and

[46] Since our SFC is informative about storage costs in the PADD3 area, we should ideally focus on the companies that are most active in that region. Alas, we do not have detailed information about where each company generates most of its profits. If anything, this limitation suggests that the forecasting power we document in this paper is a conservative estimate of the predictive power of the *scc* for the stock returns of related companies.

[47] This index includes companies such as Plains GP Holdings and Magellan Midstream Partners that are very active in the transportation and storage of crude oil. Clearly, these companies are leading firms in their line of business. For example, Plains GP Holdings is the largest storage operator at Cushing. Magellan Midstream Partners is associated with the Permian WTI storage futures contract.

[48] Before turning to the results, we emphasize that it is not a foregone conclusion that the *scc* predicts the returns on these ETFs. After all, the Alerian infrastructure index is not specific to the PADD3 region. Indeed, it captures information about (i) crude oil in the PADD3 and other regions and (ii) other energy segments, e.g., natural gas.

acyc. Our empirical results reveal that the *scc* plays a key role in the dynamics of the *basis*. Indeed, it captures most of the mean of *basis* and accounts for around 50% of its variations.

Consistent with the theory of storage, the *basis* predicts the spot return of sour crude oil. Digging deeper, we establish that the *scc* is the main conduit through which this predictive power arises. We also document the predictive power of the *scc* for important quantities, including inventory growth, the spot return on WTI crude oil, and the return on ETFs tracking companies active in the mid-stream segment of the energy value chain.

As more data become available, it would be interesting to analyze the extent to which our predictability results hold out-of-sample. Furthermore, as more storage futures contracts reach the market, it would be insightful to (i) carry out a cross-sectional analysis of the cost of storage and (ii) explore the impact of the cross-sectional differences in the storage cost on the performance of the carry strategy (Fuertes et al., 2010; Koijen et al., 2018). Finally, it would be interesting to develop theories that can quantitatively generate all our main findings. We leave all these interesting avenues for future research.

References

Alquist, R., Bauer, G. H., and de los Rios, A. D. (2014). What does the convenience yield curve tell us about the crude oil market? *Bank of Canada Working Paper*.

Armstrong, W. J., Cardella, L., and Sabah, N. (2021). Information shocks, disagreement, and drift. *Journal of Financial Economics*, 140(3):916–940.

Avino, D. E., Stancu, A., and Wese Simen, C. (2020). The predictive power of the dividend risk premium. *Forthcoming in Journal of Financial and Quantitative Analysis*.

Baker, S. D. (2021). The financialization of storable commodities. *Management Science*, 67(1):471–499.

Boons, M. and Prado, M. P. (2019). Basis-momentum. *Journal of Finance*, 74(1):239–279.

Brennan, M. J. (1958). The supply of storage. *American Economic Review*, 48(1):50–72.

Brooks, C., Prokopczuk, M., and Wu, Y. (2013). Commodity futures prices: More evidence on forecast power, risk premia and the theory of storage. *Quarterly Review of Economics and Finance*, 53(1):73–85.

Byun, S. J. (2017). Speculation in commodity futures markets, inventories and the price of crude oil. *Energy Journal*, 38(5).

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Casassus, J. and Collin-Dufresne, P. (2005). Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance*, 60(5):2283–2331.

Chincarini, L. (2020). Tracking spot oil: The elusive quest. *Journal of Commodity Markets*, 17:1–22.

Ederington, L. H., Fernando, C. S., Holland, K. V., Lee, T. K., and Linn, S. C. (2021). Dynamics of arbitrage. *Journal of Financial and Quantitative Analysis*, 56(4):1350–1380.

Fama, E. F. and French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, pages 55–73.

Fuertes, A.-M., Miffre, J., and Rallis, G. (2010). Tactical allocation in commodity futures markets: Combining momentum and term structure signals. *Journal of Banking & Finance*, 34(10):2530–2548.

Gorton, G. B., Hayashi, F., and Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance*, 17(1):35–105.

Gu, M., Kang, W., Lou, D., and Tang, K. (2020). Commodity convexity. *Working Paper*.

Hollstein, F., Prokopczuk, M., and Wese Simen, C. (2020). The conditional capital asset pricing model revisited: Evidence from high-frequency betas. *Management Science*, 66(6):2474–2494.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, 7(1):1–27.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., and Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, 127(2):197–225.

Mukherjee, A., Panayotov, G., and Shon, J. (2021). Eye in the sky: Private satellites and government macro data. *Journal of Financial Economics*, 141(1):234–254.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Prokopczuk, M., Wese Simen, C., and Wichmann, R. (2021). The natural gas announcement day puzzle. *Energy Journal*, 42(2).

Rösch, D. M., Subrahmanyam, A., and Van Dijk, M. A. (2017). The dynamics of market efficiency. *Review of Financial Studies*, 30(4):1151–1187.

Ross, S. A. (1997). Hedging long run commitments: Exercises in incomplete market pricing. *Economic Notes*, 26(2):385–420.

Rousse, O. and Sévi, B. (2019). Informed trading in the WTI oil futures market. *Energy Journal*, 40(2).

Routledge, B. R., Seppi, D. J., and Spatt, C. S. (2000). Equilibrium forward curves for commodities. *Journal of Finance*, 55(3):1297–1338.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235.

Symeonidis, L., Prokopczuk, M., Brooks, C., and Lazar, E. (2012). Futures basis, inventory and commodity price volatility: An empirical analysis. *Economic Modelling*, 29(6):2651–2663.

Szymanowska, M., De Roon, F., Nijman, T., and Van Den Goorbergh, R. (2014). An anatomy of commodity futures risk premia. *Journal of Finance*, 69(1):453–482.

Working, H. (1949). The theory of price of storage. *American Economic Review*, 39(6):1254–1262.

Figure 1: Time-Series Dynamics of the Basis

This figure depicts the time-series dynamics of the 1-month basis (*basis*). The sample period is from January 2016 until December 2019. The vertical axis shows the level of the *basis*. The horizontal axis indicates the observation date. The shaded bars indicate periods of contango, defined as periods associated with a positive *basis*.

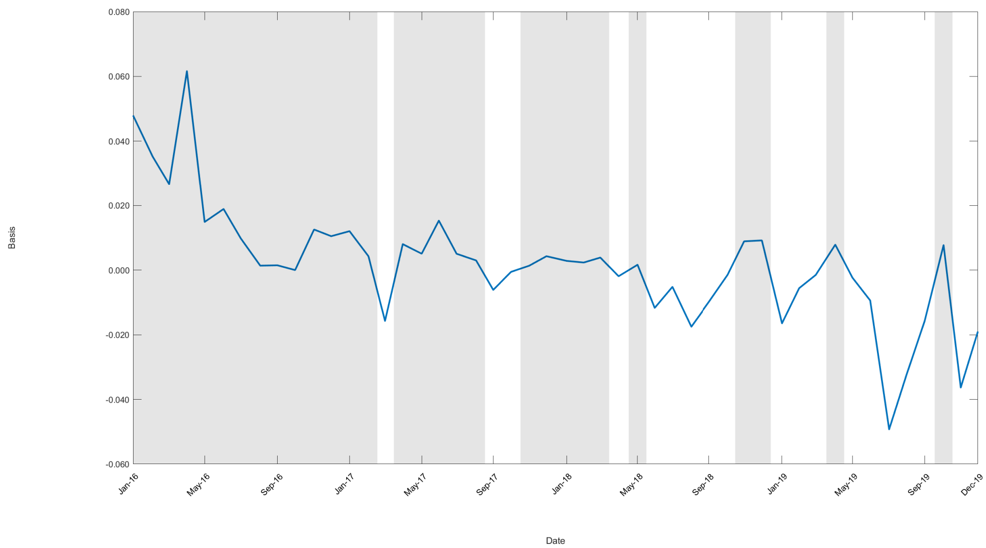

Figure 2: Time-Series Dynamics of the Storage Cost and Adjusted Convenience Channels

This figure depicts the time-series dynamics of the 1-month storage cost channel (*scc*) and the 1-month adjusted convenience yield channel (*acyc*). The sample period is from January 2016 until December 2019. The vertical axis shows the level of the *scc* (Top Plot) and *acyc* (Bottom Plot). The horizontal axis indicates the observation date. The shaded bars indicate periods of contango.

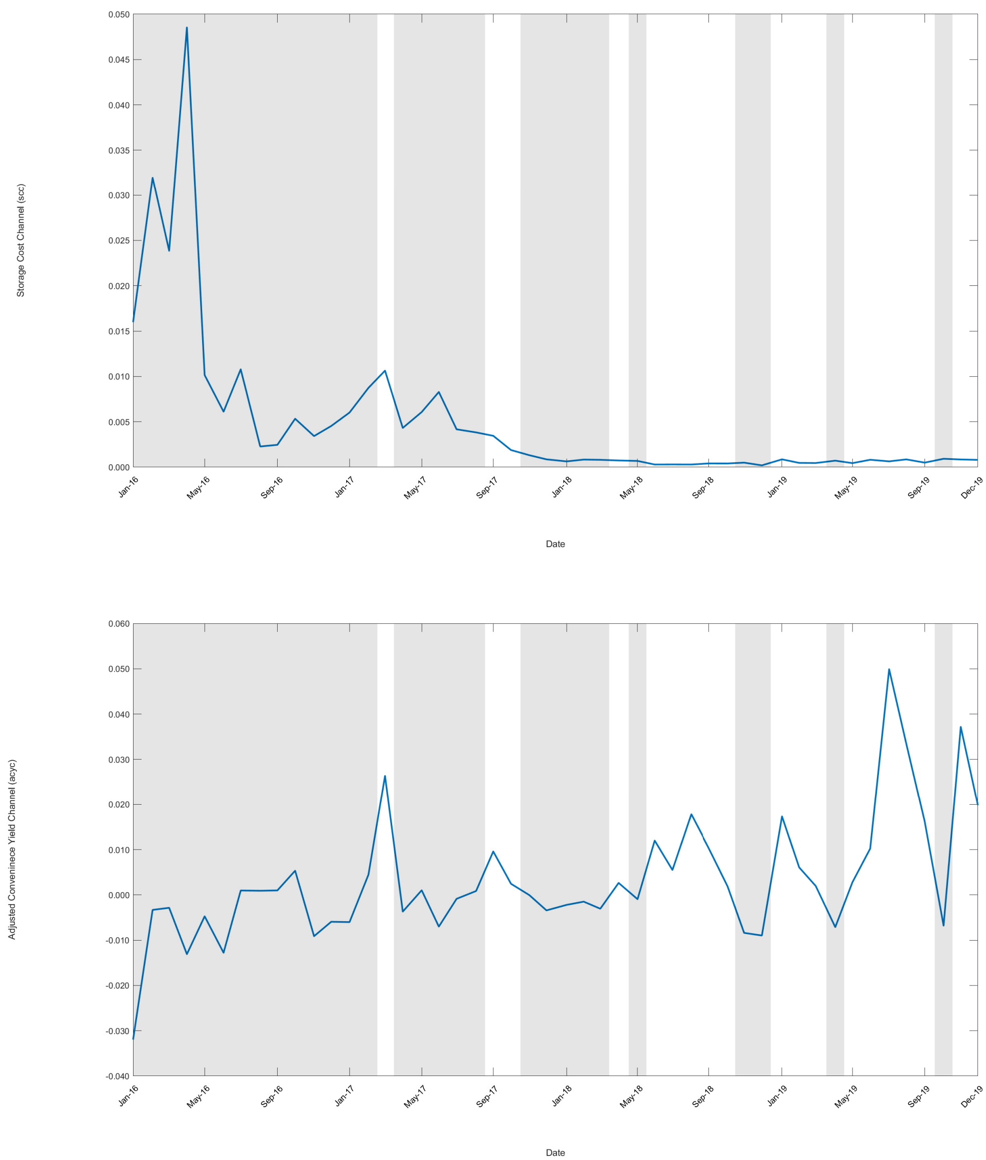

Table 1: Summary Statistics

This table presents the summary statistics of the basis ($basis$), the storage cost channel (scc), and the adjusted convenience yield channel ($acyc$). The sample period is from January 2016 until December 2019. $Mean$ is the average. $Median$ indicates the median. Std is the standard deviation. $Skew$ and $Kurt$ are the skewness and kurtosis, respectively. Min and Max point to the minimum and maximum values of the series. Finally, $AR(1)$ and $Nobs$ show the first order autocorrelation estimate as well as the total number of observations.

	$basis$	scc	$acyc$
$Mean$	0.18%	0.50%	0.32%
$Median$	0.20%	0.09%	0.09%
Std	1.87%	0.89%	1.38%
$Skew$	0.37	3.28	1.07
$Kurt$	5.40	14.64	5.51
Min	−4.93%	0.02%	−3.20%
Max	6.16%	4.85%	4.99%
$AR(1)$	0.56	0.63	0.36
$Nobs$	48	48	48

Table 2: Dissecting the Basis

This table summarizes the results of the decomposition of the basis (*basis*) into components linked to (i) the storage cost channel (*scc*) and (ii) the adjusted convenience yield channel (*acyc*). Mean shows the contribution of the variable [name in column] to the average of *basis*. Variance presents the contribution of the variable [name in column] to the variance of *basis*. Panel A shows the results based on the entire sample. Panels B and C show the results based on samples linked to observations taken during the backwardation and contango periods, respectively.

Panel A: Unconditional

	scc	*acyc*
Mean	281.05%	181.05%
Variance	45.35%	54.65%

Panel B: Backwardation

	scc	*acyc*
Mean	−9.70%	−109.70%
Variance	−3.45%	103.45%

Panel C: Contango

	scc	*acyc*
Mean	62.36%	−37.64%
Variance	76.34%	23.66%

Table 3: Predictability of Inventory Growth

This table summarizes the results of the predictability of the 1-month PADD3 inventory growth rate observed at time $t+1$ on a constant and several variables. α denotes the intercept. $I_{\text{contango},t}$ is a dummy that takes value 1 if the futures market is in contango at time t. $I_{\text{backwardation},t}$ is a dummy that takes value 1 if the futures market is in backwardation at time t. $I_{\text{spare capacity}<q_{50},t}$ and $I_{\text{spare capacity}>q_{50},t}$ are dummy variables taking values 1 if the spare capacity is below and above the median at time t, respectively. $\%\Delta scc_t$ and $\%\Delta acyc_t$ denote the growth rate in the storage cost and the adjusted convenience yield channels at time t, respectively. $\%\Delta imports_{t+1}$, $\%\Delta refinery_{t+1}$, and $\%\Delta production_{t+1}$ indicate the growth rate in the imports, refinery inputs, and production at time $t+1$, respectively. $\%\Delta Inv_t$ is the lagged inventory growth rate. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	−0.001	−0.001	−0.002				
	(−0.166)	(−0.145)	(−0.404)				
	[−0.152]	[−0.135]	[−0.402]				
$I_{\text{contango},t}$				0.003	0.002		
				(0.459)	(0.239)		
				[0.423]	[0.243]		
$I_{\text{backwardation},t}$				−0.008	−0.009		
				(−0.865)	(−1.192)		
				[−0.855]	[−1.172]		
$I_{\text{spare capacity}<q_{50},t}$						−0.010	−0.010
						(−1.274)	(−1.321)
						[−1.250]	[−1.350]
$I_{\text{spare capacity}>q_{50},t}$						0.008	0.008
						(0.942)	(0.814)
						[0.881]	[0.850]
$\%\Delta scc_t$	0.020	0.021	0.020				
	(2.802)	(2.728)	(2.157)				
	[2.638]	[2.546]	[2.056]				
$\%\Delta scc_t \times I_{\text{contango},t}$				0.042	0.042		
				(4.379)	(4.286)		
				[3.865]	[3.838]		
$\%\Delta scc_t \times I_{\text{backwardation},t}$				0.009	0.007		
				(1.877)	(1.376)		
				[1.861]	[1.454]		
$\%\Delta scc_t \times I_{\text{spare capacity}<q_{50},t}$						0.045	0.048
						(3.641)	(3.541)
						[3.486]	[3.436]
$\%\Delta scc_t \times I_{\text{spare capacity}>q_{50},t}$						0.007	0.006
						(1.440)	(0.929)
						[1.361]	[0.951]
$\%\Delta acyc_t$			−0.001		−0.001		−0.001
			(−3.951)		(−3.288)		(−3.527)
			[−4.103]		[−3.371]		[−3.668]
$\%\Delta imports_{t+1}$			0.015		0.011		−0.011
			(0.499)		(0.379)		(−0.377)
			[0.522]		[0.407]		[−0.390]
$\%\Delta refinery_{t+1}$			−0.179		−0.204		−0.180
			(−1.864)		(−2.334)		(−2.060)
			[−1.960]		[−2.400]		[−2.137]
$\%\Delta production_{t+1}$			0.331		0.283		0.180
			(2.751)		(2.358)		(1.373)
			[2.628]		[2.211]		[1.370]
$\%\Delta Inv_t$		−0.060	0.042	−0.138	−0.033	−0.016	0.066
		(−0.491)	(0.290)	(−1.001)	(−0.215)	(−0.143)	(0.539)
		[−0.463]	[0.284]	[−0.971]	[−0.216]	[−0.132]	[0.519]
Adj R^2	7.07%	5.29%	12.97%	9.11%	17.57%	10.59%	18.69%
$Nobs$	46	46	45	46	45	46	45

Table 4: Spot Return Predictability: LOOP Gulf Coast Crude Oil

This table summarizes the results of predictive regressions of the 1-month spot return of the Gulf Coast sour crude oil on a constant and the lagged variable [name in row]. α denotes the intercept. *basis* denotes the basis. *scc* denotes the storage cost channel. *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *relbasis* is the relative basis signal. *mom* and *basmom* denote the momentum and basis momentum signals, respectively. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	0.022	0.005	0.034	0.012	0.016	0.039	0.024
	(1.814)	(0.365)	(1.978)	(0.795)	(0.923)	(2.158)	(1.160)
	[1.747]	[0.355]	[2.009]	[0.762]	[0.934]	[2.241]	[1.204]
basis	2.241						
	(2.382)						
	[2.379]						
scc		4.475		3.800	3.034		2.647
		(3.014)		(3.334)	(2.958)		(2.805)
		[2.858]		[2.981]	[2.605]		[2.767]
acyc			−2.196	−1.420		−2.605	−2.335
			(−1.654)	(−1.292)		(−1.585)	(−1.445)
			[−1.651]	[−1.265]		[−1.631]	[−1.487]
relbasis					−0.009	1.710	1.814
					(−0.014)	(1.102)	(1.207)
					[−0.013]	[1.138]	[1.227]
mom					−0.079	−0.105	−0.067
					(−1.458)	(−2.295)	(−1.354)
					[−1.459]	[−2.306]	[−1.389]
basmom					−0.050	−0.002	−0.013
					(−0.152)	(−0.006)	(−0.038)
					[−0.149]	[−0.006]	[−0.037]
Adj R^2	15.04%	13.87%	6.80%	15.40%	12.16%	13.25%	15.33%
$Nobs$	47	47	47	47	47	47	47

Table 5: Dissecting Convexity

This table summarizes the full sample results of the decomposition of the convexity (*convexity*) into components linked to (i) the storage cost channel ($scc^{convexity}$) and (ii) the adjusted convenience yield channel ($acyc^{convexity}$). Mean shows the contribution of the variable [name in column] to the average of *convexity*. Variance presents the contribution of the variable [name in column] to the variance of *convexity*.

	scc	*acyc*
Mean	289.93%	189.93%
Variance	17.69%	82.31%

Table 6: Spot Return Predictability: WTI Crude Oil

This table summarizes the results of predictive regressions of the 1-month WTI spot return on a constant and the lagged variable [name in row]. α denotes the intercept. *basis* denotes the basis. *scc* denotes the storage cost channel. *acyc* indicates the adjusted convenience yield channel. *relbasis* is the relative basis signal. *mom* and *basmom* denote the momentum and basis momentum signals, respectively. Note that all explanatory variables are computed using information from the Gulf Coast sour crude oil market. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. *Nobs* shows the total number of observations.

α	0.019	0.002	0.024	0.001	0.008	0.029	0.011
	(1.597)	(0.155)	(1.545)	(0.096)	(0.537)	(1.806)	(0.649)
	[1.580]	[0.152]	[1.619]	[0.093]	[0.543]	[1.919]	[0.665]
basis	1.295						
	(1.654)						
	[1.714]						
scc		3.920		3.976	3.153		3.023
		(2.739)		(2.903)	(2.679)		(2.675)
		[2.481]		[2.524]	[2.351]		[2.407]
acyc			−0.695	0.117		−1.096	−0.788
			(−0.806)	(0.186)		(−1.019)	(−0.790)
			[−0.812]	[0.176]		[−1.055]	[−0.814]
relbasis					1.107	1.604	1.722
					(1.661)	(1.329)	(1.406)
					[1.613]	[1.334]	[1.429]
mom					−0.059	−0.099	−0.055
					(−1.403)	(−2.221)	(−1.325)
					[−1.435]	[−2.309]	[−1.363]
basmom					0.018	0.043	0.030
					(0.058)	(0.134)	(0.096)
					[0.057]	[0.134]	[0.096]
Adj R^2	5.08%	13.43%	−1.08%	11.49%	12.46%	6.44%	11.09%
Nobs	47	47	47	47	47	47	47

Table 7: Spot Return Predictability: Mid-Stream ETFs

This table summarizes the results of predictive regressions of the 1-month return on the ETF [name in column] on a constant and the lagged variable [name in row]. α denotes the intercept. scc is the storage cost channel. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). R^2 indicates the R-squared of the regression. Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

	ENFR	**EINC**	**AMLP**
α	0.001	−0.003	−0.003
	(0.130)	(−0.424)	(−0.425)
	[0.122]	[−0.367]	[−0.378]
scc	2.365	2.628	2.405
	(2.511)	(3.468)	(2.592)
	[2.376]	[3.086]	[2.617]
Adj R^2	13.78%	16.43%	9.20%
$Nobs$	47	47	47

The Dynamics of Storage Costs

Online Appendix

JEL classification: G13, G14, G17

Keywords: Basis, Futures, Inventory, Predictability, Storage Costs

Table A1: Logarithmic Growth of Inventory

This table reports the results of the predictability of the (log) growth rate of the 1-month PADD3 inventory observed at time $t+1$ on a constant and several variables. α denotes the intercept. $I_{\text{contango},t}$ is a dummy that takes value 1 if the futures market is in contango at time t. $I_{\text{backwardation},t}$ is a dummy that takes value 1 if the futures market is in backwardation at time t. $I_{\text{spare capacity}<q_{50},t}$ and $I_{\text{spare capacity}>q_{50},t}$ are dummy variables taking values 1 if the spare capacity is below and above the median at time t, respectively. $\%\Delta scc_t$ and $\%\Delta acyc_t$ denote the log growth rate in the storage cost channel and the simple growth of the adjusted convenience yield at t, respectively. $\%\Delta imports_{t+1}$, $\%\Delta refinery_{t+1}$, and $\%\Delta production_{t+1}$ indicate the (log) growth rate in the imports, the refinery inputs, and the production at time $t+1$, respectively. $\%\Delta Inv_t$ is the lagged (log) inventory growth rate. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	0.001	0.002	0.000				
	(0.263)	(0.271)	(−0.013)				
	[0.240]	[0.252]	[−0.013]				
$I_{\text{contango},t}$				0.008	0.006		
				(1.136)	(0.831)		
				[1.042]	[0.849]		
$I_{\text{backwardation},t}$				−0.008	−0.009		
				(−0.854)	(−1.283)		
				[−0.844]	[−1.237]		
$I_{\text{spare capacity}<q_{50},t}$						−0.005	−0.005
						(−0.616)	(−0.607)
						[−0.616]	[−0.630]
$I_{\text{spare capacity}>q_{50},t}$						0.008	0.006
						(1.021)	(0.705)
						[0.949]	[0.714]
$\%\Delta scc_{t-1}$	0.030	0.031	0.032				
	(4.462)	(4.287)	(4.059)				
	[4.089]	[3.968]	[3.820]				
$\%\Delta scc_{t-1} \times I_{\text{contango},t}$				0.038	0.037		
				(4.228)	(4.312)		
				[4.066]	[4.099]		
$\%\Delta scc_{t-1} \times I_{\text{backwardation},t}$				0.017	0.021		
				(1.481)	(1.753)		
				[1.509]	[1.759]		
$\%\Delta scc_{t-1} \times I_{\text{spare capacity}<q_{50},t}$						0.037	0.038
						(3.139)	(3.056)
						[2.984]	[2.990]
$\%\Delta scc_{t-1} \times I_{\text{spare capacity}>q_{50},t}$						0.020	0.021
						(2.061)	(1.961)
						[1.996]	[1.881]
$\%\Delta acyc_t$			−0.001		−0.001		−0.001
			(−5.376)		(−4.531)		(−4.392)
			[−5.429]		[−4.631]		[−4.529]
$\%\Delta imports_{t+1}$			0.009		0.008		−0.001
			(0.325)		(0.262)		(−0.032)
			[0.340]		[0.284]		[−0.033]
$\%\Delta refinery_{t+1}$			−0.176		−0.172		−0.174
			(−2.191)		(−2.124)		(−2.208)
			[−2.251]		[−2.159]		[−2.249]
$\%\Delta production_{t+1}$			0.359		0.321		0.269
			(3.135)		(2.806)		(2.011)
			[2.971]		[2.637]		[1.985]
$\%\Delta Inv_t$		−0.089	0.023	−0.145	−0.036	−0.050	0.040
		(−0.752)	(0.167)	(−1.112)	(−0.241)	(−0.472)	(0.316)
		[−0.722]	[0.164]	[−1.074]	[−0.240]	[−0.442]	[0.308]
Adj R^2	0.129	0.117	0.225	0.118	0.219	0.109	0.208
$Nobs$	46	46	46	46	46	46	46

Table A2: Forecasting the Inventory Changes

This table summarizes the results of the predictability of the 1-month PADD3 inventory change observed at time $t+1$ on a constant and several variables. α denotes the intercept. $I_{\text{contango},t}$ is a dummy that takes value 1 if the futures market is in contango at time t. $I_{\text{backwardation},t}$ is a dummy that takes value 1 if the futures market is in backwardation at time t. $I_{\text{spare capacity}<q_{50},t}$ and $I_{\text{spare capacity}>q_{50},t}$ are dummy variables taking values 1 if the spare capacity is below and above the median at time t, respectively. Δscc_t and $\Delta acyc_t$ denote the change in the storage cost channel and the change in the adjusted convenience yield channel at t, respectively. $\Delta imports_{t+1}$, $\Delta refinery_{t+1}$, and $\Delta production_{t+1}$ indicate the change in the imports, the refinery inputs, and the production at time $t+1$, respectively. $\%\Delta Inv_t$ is the lagged change in the inventory level. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	−0.002	−0.002	−0.003				
	(−0.291)	(−0.265)	(−0.538)				
	[−0.265]	[−0.246]	[−0.534]				
$I_{\text{contango},t}$				0.002	0.000		
				(0.284)	(0.066)		
				[0.257]	[0.067]		
$I_{\text{backwardation},t}$				−0.009	−0.009		
				(−0.887)	(−1.240)		
				[−0.877]	[−1.219]		
$I_{\text{spare capacity}<q_{50},t}$						−0.012	−0.012
						(−1.641)	(−1.752)
						[−1.627]	[−1.804]
$I_{\text{spare capacity}>q_{50},t}$						0.008	0.008
						(0.914)	(0.831)
						[0.858]	[0.869]
Δscc_{t-1}	0.020	0.020	0.019				
	(2.578)	(2.525)	(1.979)				
	[2.484]	[2.412]	[1.923]				
$\Delta scc_{t-1} \times I_{\text{contango},t}$				0.041	0.042		
				(5.967)	(6.351)		
				[5.367]	[5.630]		
$\Delta scc_{t-1} \times I_{\text{backwardation},t}$				0.008	0.007		
				(2.745)	(1.540)		
				[2.609]	[1.615]		
$\Delta scc_{t-1} \times I_{\text{spare capacity}<q_{50},t}$						0.043	0.047
						(5.180)	(5.139)
						[4.983]	[5.015]
$\Delta scc_{t-1} \times I_{\text{spare capacity}>q_{50},t}$						0.007	0.005
						(1.751)	(0.984)
						[1.644]	[1.003]
$\Delta acyc_t$			−0.001		−0.001		−0.001
			(−4.160)		(−3.865)		(−3.905)
			[−4.317]		[−3.947]		[−4.079]
$\Delta imports_{t+1}$			0.011		0.004		−0.016
			(0.365)		(0.155)		(−0.561)
			[0.380]		[0.165]		[−0.578]
$\Delta refinery_{t+1}$			−0.173		−0.202		−0.181
			(−1.792)		(−2.357)		(−2.090)
			[−1.883]		[−2.414]		[−2.146]
$\Delta production_{t+1}$			0.333		0.282		0.179
			(2.758)		(2.359)		(1.403)
			[2.646]		[2.230]		[1.404]
$\%\Delta Inv_t$		−0.059	0.043	−0.139	−0.036	−0.026	0.057
		(−0.505)	(0.304)	(−1.061)	(−0.241)	(−0.255)	(0.490)
		[−0.470]	[0.295]	[−1.023]	[−0.241]	[−0.232]	[0.468]
Adj R^2	0.090	0.072	0.150	0.128	0.222	0.143	0.241
$Nobs$	46	46	46	46	46	46	46

Table A3: Spot (Log) Return Predictability: LOOP Gulf Coast Crude Oil

This table summarizes the results of predictive regressions of the 1-month spot logarithmic return of the Gulf Coast sour crude oil on a constant and the lagged variable [name in row]. α denotes the intercept. *basis* is the basis. *scc* denotes the storage cost channel. *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *relbasis* is the relative basis signal, computed as in Gu et al. (2020). *mom* is the momentum signal. Specifically, this is the return of the first nearby futures contract of the LOOP Gulf Coast crude oil over the most recent 12-month period. *basmom* is the difference between the momentum signals computed using the first and second nearbys of the LOOP Gulf Coast crude oil market. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	0.018	0.001	0.028	0.008	0.011	0.033	0.018
	(1.474)	(0.110)	(1.731)	(0.501)	(0.678)	(1.918)	(0.922)
	[1.415]	[0.107]	[1.751]	[0.481]	[0.684]	[1.980]	[0.952]
basis	1.999						
	(2.435)						
	[2.433]						
scc		4.118		3.558	2.839		2.514
		(3.157)		(3.401)	(2.987)		(2.773)
		[2.988]		[3.046]	[2.653]		[2.734]
acyc			−1.905	−1.178		−2.216	−1.960
			(−1.638)	(−1.213)		(−1.515)	(−1.371)
			[−1.631]	[−1.183]		[−1.562]	[−1.412]
relbasis					0.025	1.457	1.555
					(0.040)	(1.037)	(1.148)
					[0.037]	[1.064]	[1.157]
mom					−0.071	−0.097	−0.060
					(−1.421)	(−2.235)	(−1.308)
					[−1.419]	[−2.235]	[−1.332]
basmom					−0.032	0.010	−0.001
					(−0.094)	(0.027)	(−0.002)
					[−0.091]	[0.026]	[−0.002]
Adj R^2	12.54%	12.43%	5.07%	13.02%	9.98%	9.94%	11.81%
$Nobs$	47	47	47	47	47	47	47

Table A4: Inventory Growth Predictability Assuming Release on Friday

This table summarizes the results of the predictability of the 1-month PADD3 inventory growth rate observed at time $t+1$ on a constant and several variables. We assume that the inventory data is known to market participants on the measurement date, i.e. on Friday, rather than on the publication date, i.e. the following Wednesday. α denotes the intercept. $I_{\text{contango},t}$ is a dummy that takes value 1 if the futures market is in contango at time t. $I_{\text{backwardation},t}$ is a dummy that takes value 1 if the futures market is in backwardation at time t. $I_{\text{spare capacity}<q_{50},t}$ and $I_{\text{spare capacity}>q_{50},t}$ are dummy variables taking values 1 if the spare capacity is below and above the median at time t, respectively. $\%\Delta scc_t$ and $\%\Delta acyc_t$ denote the growth rate in the storage cost channel and the adjusted convenience yield channel at t, respectively. $\%\Delta imports_{t+1}$, $\%\Delta refinery_{t+1}$, and $\%\Delta production_{t+1}$ indicate the growth rate in the imports, the refinery inputs and the production at time $t+1$, respectively. $\%\Delta Inv_t$ is the lagged inventory growth rate. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	−0.001	−0.001	−0.006				
	(−0.179)	(−0.166)	(−1.030)				
	[−0.167]	[−0.157]	[−1.001]				
$I_{\text{contango},t}$				0.003	−0.002		
				(0.390)	(−0.288)		
				[0.362]	[−0.284]		
$I_{\text{backwardation},t}$				−0.008	−0.012		
				(−0.735)	(−1.449)		
				[−0.741]	[−1.442]		
$I_{\text{spare capacity}<q_{50},t}$						−0.011	−0.012
						(−1.570)	(−1.772)
						[−1.465]	[−1.678]
$I_{\text{spare capacity}>q_{50},t}$						0.009	0.003
						(0.987)	(0.318)
						[0.962]	[0.329]
$\%\Delta scc_t$	0.018	0.018	0.021				
	(1.979)	(1.986)	(2.066)				
	[1.917]	[1.888]	[2.007]				
$\%\Delta scc_t \times I_{\text{contango},t}$				0.044	0.043		
				(3.858)	(3.613)		
				[3.434]	[3.431]		
$\%\Delta scc_t \times I_{\text{backwardation},t}$				0.004	0.007		
				(0.550)	(1.226)		
				[0.567]	[1.289]		
$\%\Delta scc_t \times I_{\text{spare capacity}<q_{50},t}$						0.049	0.050
						(3.380)	(3.123)
						[3.248]	[3.095]
$\%\Delta scc_t \times I_{\text{spare capacity}>q_{50},t}$						0.001	0.003
						(0.174)	(0.630)
						[0.173]	[0.635]
$\%\Delta acyc_t$			0.000		0.000		−0.001
			(−1.956)		(−1.639)		(−2.077)
			[−1.989]		[−1.695]		[−2.120]
$\%\Delta imports_{t+1}$			−0.033		−0.022		−0.038
			(−0.909)		(−0.554)		(−1.232)
			[−0.914]		[−0.581]		[−1.220]
$\%\Delta refinery_{t+1}$			−0.112		−0.139		−0.119
			(−1.285)		(−1.663)		(−1.499)
			[−1.256]		[−1.664]		[−1.458]
$\%\Delta production_{t+1}$			0.720		0.685		0.553
			(3.980)		(4.403)		(2.620)
			[3.574]		[3.765]		[2.355]
$\%\Delta Inv_t$		−0.046	0.063	−0.120	−0.006	0.006	0.105
		(−0.388)	(0.439)	(−0.893)	(−0.039)	(0.052)	(0.786)
		[−0.361]	[0.417]	[−0.876]	[−0.039]	[0.048]	[0.750]
Adj R^2	0.046	0.026	0.132	0.088	0.187	0.121	0.215
$Nobs$	46	46	46	46	46	46	46

Table A5: Summary Statistics (Treasury Rate)

This table presents the summary statistics of the basis (*basis*), the storage cost channel (*scc*), and the adjusted convenience yield channel (*acyc*). The *scc* is estimated using the Treasury rates (instead of Libor rates). The sample period is from January 2016 until December 2019. *Mean* is the average. *Median* indicates the median. *Std* is the standard deviation of the time-series. *Skew* and *Kurt* relate to the skewness and kurtosis, respectively. *Min* and *Max* are the minimum and maximum values of the relevant series. Finally, $AR(1)$ and *Nobs* show the first order autocorrelation estimate and the total number of observations, respectively. We compute the statistics separately for the time-series of each variable [name in column]. Panel A uses information from the full sample. Panel B is based on the periods of backwardation, defined as periods where the basis is negative. Finally, Panel C uses data from the contango period only.

Panel A: Unconditional

	basis	*scc*	*acyc*
Mean	0.18%	0.50%	0.32%
Median	0.20%	0.09%	0.09%
Std	1.87%	0.89%	1.38%
Skew	0.37	3.28	1.07
Kurt	5.40	14.64	5.51
Min	−4.93%	0.02%	−3.20%
Max	6.16%	4.85%	4.99%
$AR(1)$	0.56	0.63	0.36
Nobs	48	48	48

Panel B: Backwardation

	basis	*scc*	*acyc*
Mean	−1.36%	0.13%	1.49%
Median	−0.99%	0.06%	1.03%
Std	1.33%	0.24%	1.35%
Skew	−1.32	3.46	1.16
Kurt	4.05	13.97	3.58
Min	−4.93%	0.03%	0.19%
Max	−0.06%	1.06%	4.99%
$AR(1)$	0.39	0.29	0.36
Nobs	19	19	19

Panel C: Contango

	basis	*scc*	*acyc*
Mean	1.19%	0.74%	−0.45%
Median	0.78%	0.42%	−0.33%
Std	1.44%	1.07%	0.70%
Skew	2.16	2.54	−2.09
Kurt	7.19	9.31	9.34
Min	0.00%	0.02%	−3.20%
Max	6.16%	4.85%	0.53%
$AR(1)$	0.51	0.58	0.04
Nobs	29	29	29

Table A6: Dissecting the Basis (Treasury Rate)

This table summarizes the results of the decomposition of the basis (*basis*) into components linked to (i) the storage cost channel (*scc*) and (ii) the adjusted convenience yield channel (*acyc*). The *scc* is estimated using the Treasury rates (instead of Libor rates). Mean shows the contribution of the variable [name in column] to the average of *basis*. Variance presents the contribution of the variable [name in column] to the variance of *basis*. Panel A shows the results when the decomposition is carried out using all sample information. Panels B and C show the results based on the sample of observations linked to backwardation and contango periods, respectively.

Panel A: Unconditional

	scc	*acyc*
Mean	281.00%	181.00%
Variance	45.35%	54.65%

Panel B: Backwardation

	scc	*acyc*
Mean	−9.70%	−109.70%
Variance	−3.45%	103.45%

Panel C: Contango

	scc	*acyc*
Mean	62.35%	−37.65%
Variance	76.33%	23.67%

Table A7: Inventory Growth Predictability (Treasury Rate)

This table summarizes the results of the predictability of the 1-month PADD3 inventory growth rate observed at time $t+1$ on a constant and several variables. α is the intercept. $I_{\text{contango},t}$ is a dummy that takes value 1 if the futures market is in contango at time t. $I_{\text{backwardation},t}$ is a dummy that takes value 1 if the futures market is in backwardation at time t. $I_{\text{spare capacity}<q_{50},t}$ and $I_{\text{spare capacity}>q_{50},t}$ are dummy variables taking values 1 if the spare capacity is below and above the median at time t, respectively. The scc is estimated using the Treasury rates (instead of Libor rates). $\%\Delta scc_t$ and $\%\Delta acyc_t$ denote the growth rate in the storage cost and the adjusted convenience yield channels at t, respectively. $\%\Delta imports_{t+1}$, $\%\Delta refinery_{t+1}$, and $\%\Delta production_{t+1}$ indicate the growth rate in the imports, the refinery inputs, and the production at time $t+1$, respectively. $\%\Delta Inv_t$ is the lagged inventory growth rate. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	−0.001	−0.001	−0.002				
	(−0.166)	(−0.145)	(−0.405)				
	[−0.152]	[−0.135]	[−0.403]				
$I_{\text{contango},t}$				0.003	0.002		
				(0.459)	(0.238)		
				[0.423]	[0.243]		
$I_{\text{backwardation},t}$				−0.008	−0.009		
				(−0.865)	(−1.192)		
				[−0.855]	[−1.172]		
$I_{\text{spare capacity}<q_{50},t}$						−0.010	−0.010
						(−1.274)	(−1.321)
						[−1.250]	[−1.350]
$I_{\text{spare capacity}>q_{50},t}$						0.008	0.008
						(0.941)	(0.814)
						[0.881]	[0.850]
$\%\Delta scc_t$	0.020	0.021	0.020				
	(2.802)	(2.728)	(2.157)				
	[2.638]	[2.546]	[2.056]				
$\%\Delta scc_t \times I_{\text{contango},t}$				0.042	0.042		
				(4.380)	(4.287)		
				[3.865]	[3.839]		
$\%\Delta scc_t \times I_{\text{backwardation},t}$				0.009	0.007		
				(1.878)	(1.376)		
				[1.862]	[1.454]		
$\%\Delta scc_t \times I_{\text{spare capacity}<q_{50},t}$						0.045	0.048
						(3.642)	(3.541)
						[3.487]	[3.436]
$\%\Delta scc_t \times I_{\text{spare capacity}>q_{50},t}$						0.007	0.006
						(1.441)	(0.930)
						[1.362]	[0.951]
$\%\Delta acyc_t$			−0.001		−0.001		−0.001
			(−3.936)		(−3.275)		(−3.516)
			[−4.085]		[−3.357]		[−3.655]
$\%\Delta imports_{t+1}$			0.015		0.011		−0.011
			(0.499)		(0.379)		(−0.378)
			[0.522]		[0.407]		[−0.391]
$\%\Delta refinery_{t+1}$			−0.179		−0.204		−0.180
			(−1.864)		(−2.334)		(−2.060)
			[−1.960]		[−2.400]		[−2.136]
$\%\Delta production_{t+1}$			0.331		0.283		0.179
			(2.750)		(2.357)		(1.372)
			[2.627]		[2.210]		[1.369]
$\%\Delta Inv_t$		−0.060	0.042	−0.138	−0.033	−0.016	0.066
		(−0.491)	(0.289)	(−1.001)	(−0.215)	(−0.143)	(0.538)
		[−0.463]	[0.283]	[−0.971]	[−0.216]	[−0.132]	[0.518]
Adj R^2	0.071	0.053	0.130	0.091	0.176	0.106	0.187
$Nobs$	46	46	46	46	46	46	46

Table A8: Spot Return Predictability: LOOP Gulf Coast Crude Oil (Treasury Rate)

This table summarizes the results of regressions of the 1-month Gulf Coast crude oil spot return on a constant and the lagged variable [name in row]. α denotes the intercept. *basis* denotes the basis. *scc* denotes the storage cost channel. The *scc* is estimated using the Treasury rates (instead of Libor rates). *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *relbasis* is the relative basis signal. *mom* and *basmom* denote the momentum and basis momentum signals, respectively. *basis*,*relbasis*, *mom*, and *basmom* are computed using information from the Gulf Coast sour crude oil market. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	0.022	0.005	0.034	0.012	0.016	0.039	0.024
	(1.814)	(0.365)	(1.978)	(0.795)	(0.923)	(2.158)	(1.160)
	[1.747]	[0.355]	[2.009]	[0.762]	[0.934]	[2.241]	[1.204]
basis	2.241						
	(2.382)						
	[2.379]						
scc		4.476		3.800	3.035		2.647
		(3.015)		(3.334)	(2.958)		(2.804)
		[2.858]		[2.981]	[2.605]		[2.767]
acyc			−2.197	−1.420		−2.605	−2.335
			(−1.654)	(−1.292)		(−1.585)	(−1.445)
			[−1.652]	[−1.265]		[−1.631]	[−1.487]
relbasis					−0.009	1.710	1.814
					(−0.014)	(1.102)	(1.207)
					[−0.013]	[1.138]	[1.227]
mom					−0.079	−0.105	−0.067
					(−1.458)	(−2.295)	(−1.354)
					[−1.459]	[−2.306]	[−1.389]
basmom					−0.050	−0.002	−0.013
					(−0.152)	(−0.006)	(−0.038)
					[−0.149]	[−0.006]	[−0.037]
Adj R^2	15.04%	13.87%	6.80%	15.40%	12.16%	13.25%	15.33%
$Nobs$	47	47	47	47	47	47	47

Table A9: Decomposition of the Basis: Low and High Spare Capacity

This table summarizes the results of the decomposition of the basis (*basis*) into components linked to (i) the storage cost channel (*scc*) and (ii) the adjusted convenience yield channel (*acyc*). Mean shows the contribution of the variable [name in column] to the average of *basis*. Variance presents the contribution of the variable [name in column] to the variance of *basis*. Panel A shows the results based on periods of low spare capacity. Panel B presents the results based on the sample linked to observations taken during times of high spare capacity.

Panel A. Low Spare Capacity

	Storage Channel	Adjusted Convenience Channel
Mean	83.16%	−16.84%
Variance	64.70%	35.30%

Panel B. High Spare Capacity

	Storage Channel	Adjusted Convenience Channel
Mean	−9.16%	−109.16%
Variance	0.08%	99.92%

Table A10: Spot Return Predictability: WTI Crude Oil with Market Specific Controls

This table summarizes the results of the regression of the 1-month WTI spot return on a constant and the lagged variable [name in row]. α denotes the intercept. *basis* denotes the basis. *scc* denotes the storage cost channel. *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *acyc* indicates the adjusted convenience yield channel. We compute this quantity as the difference between *basis* and *scc*. *relbasis* is the relative basis signal. *mom* and *basmom* denote the momentum and basis momentum signals, respectively. Note that *relbasis*, *mom*, and *basmom* are computed using information from the WTI crude oil market. We report the Newey and West (1987) t-statistic computed with 2 lags in parentheses. The figures in square brackets indicate the bootstrapped t-statistic based on the wild bootstrap of Shao (2010). Adj R^2 is the adjusted R-squared of the model. $Nobs$ shows the total number of observations.

α	0.000	0.002	0.024	0.001	0.007	0.013	0.005
	(−0.020)	(0.155)	(1.545)	(0.096)	(0.423)	(0.701)	(0.294)
	[−0.020]	[0.152]	[1.619]	[0.093]	[0.417]	[0.690]	[0.289]
basis	2.592						
	(4.155)						
	[4.219]						
scc		3.920		3.976	2.421		2.480
		(2.739)		(2.903)	(1.786)		(1.835)
		[2.481]		[2.524]	[1.531]		[1.561]
acyc			−0.695	0.117		0.150	0.306
			(−0.806)	(0.186)		(0.214)	(0.468)
			[−0.812]	[0.176]		[0.226]	[0.489]
relbasis					−0.690	−2.338	−0.943
					(−0.342)	(−1.197)	(−0.469)
					[−0.314]	[−1.163]	[−0.433]
mom					−0.086	−0.106	−0.085
					(−1.707)	(−1.884)	(−1.662)
					[−1.737]	[−1.909]	[−1.695]
basmom					0.138	0.094	0.137
					(0.248)	(0.161)	(0.245)
					[0.266]	[0.171]	[0.263]
Adj R^2	21.53%	13.43%	−1.08%	11.49%	13.19%	9.29%	11.27%
$Nobs$	47	47	47	47	47	47	47