Relative Basis
and Risk Premia in Commodity Futures Markets

Abstract

We propose a novel measure of the basis, namely the relative basis, as a more precise measure of the convenience yield on commodity markets. The relative basis is defined as the difference between the traditional basis, which is based on the first- and second-nearby futures prices, and a similarly-defined longer-term basis that is based on the second- and third-nearby futures contracts. We argue that the relative basis is a better proxy for the convenience yield because it excludes components in the traditional basis that are related to storage and financing costs. We find that the relative basis exhibits significant return predictability on commodity futures markets — it can subsume the traditional basis in forecasting commodity futures returns in both return-predicting regressions and factor portfolio tests. Our empirical findings are robust to a variety of robustness tests and provide strong support to the Theory of Storage on commodity markets.

Keywords:
relative basis, commodity futures, risk premia, convenience yield, theory of storage

1. Introduction

The commodity futures market is of increasing importance to commodity consumers, exporting nations, as well as investors that view commodities as an integral part of their portfolios (Gorton and Rouwenhorst, 2006; Tang and Xiong, 2012). The return dynamics of commodity futures have therefore attracted significant attention from both researchers and regulators in recent years. One of the most important findings in the related literature is that commodity futures with a higher basis (i.e., those with a downward-sloping futures curve) earn significantly higher expected returns than commodity futures with a lower basis (Fama and French 1987; Gorton, Hayashi, and Rouwenhorst, 2013).

One interpretation of this empirical regularity is that the slope of the futures curve is closely tied to the demand and supply of a commodity — an idea at the core of the classic Theory of Storage. Specifically, the Theory of Storage maintains that there exists an implicit but important benefit — namely the "convenience yield" — that accrues to the owners of the physical commodity who can use the commodity for immediate production or consumption but not to the owners of the commodity futures contracts. This benefit is greatly amplified when the commodity is in "short" supply—that is, when the demand for the commodity significantly outstrips the supply.[1]

Gorton, Hayashi, and Rouwenhorst (2013) empirically show that inventories (an indicator of supply relative to demand) forecast commodity futures returns – lower inventories are associated with substantially higher futures returns. To the extent that the basis reflects the abundance (or the lack) of physical inventories, it can help forecast commodity futures returns. An obvious advantage of using the basis, instead of physical inventories, to forecast futures returns is that the basis can be measured in real time, whereas inventory data are often difficult to be obtained and/or only available with severe delays.

[1] As shown in Brennan (1958) and Pindyck (1994), the convenience yield is a decreasing and convex function of the inventory.

As we show in the paper, although the basis by itself can forecast commodity futures returns, its return predictability disappears when other variables (e.g., price momentum and basis momentum) are included. One explanation is that the basis is determined not only by the convenience yield but also by other commodity-specific characteristics, most notably the "cost of carry," which includes both financing costs and storage costs. There is likely substantial and persistent variation in storage costs across commodities due to the differences in their physical characteristics (e.g., natural gas vs. gold). As such, the basis measure used in the extant literature, which will be named as the traditional basis thereafter in this paper, is a noisy proxy for the convenience yield with many confounding factors.

We propose a simple yet effective method to isolate the convenience yield from the confounding factors by taking the difference between the traditional basis and a similarly defined longer-term basis. More specifically, our "relative basis" measure is the time-scaled difference between the prices of the first-nearby futures contract ($F_t(T1)$) and the second-nearby contract ($F_t(T2)$) minus that between the prices of the second-nearby ($F_t(T2)$) and the third-nearby contract ($F_t(T3)$). Since commodity-specific characteristics that determine storage and financing costs are persistent over time (so little change between the near and distant future), taking the difference between the short-term and long-term basis provides a cleaner measure of temporary shocks to commodity physical inventory.[2]

We then take our novel measure of relative basis to the data and provide strong evidence that it is a cleaner measure of a commodity's convenience yield than the traditional basis. To start, we show that the relative basis has much more time-variation than the traditional basis: the first-order autocorrelation in the former is 0.33, whereas that in the latter is 0.68. Moreover, the second- to the sixth-order autocorrelations of the traditional basis range between 0.3 to 0.5, while these higher-order autocorrelations for the relative basis measure are basically zero. We also find that the relative basis is more closely related to the scarcity of inventories (measured by the

[2] As shown in Gorton, Hayashi, and Rouwenhorst (2013), the inventory of a commodity is mean-reverting, so is the convenience yield.

reciprocal of inventories) than the traditional basis. Both observations corroborate the view that purging out the confounding factors in the traditional basis helps sharpen the measure.

Next, we turn to asset pricing implications of the relative basis in commodity futures markets. We find that while both the relative basis and traditional basis predict next-month futures returns in univariate regressions, the traditional basis loses its return-predictive power when other control variables are included. In contrast, the return predictive pattern of the relative basis is virtually unchanged with the addition of the traditional basis and other controls. In terms of economic magnitudes, in portfolio sorts, commodity futures in the top tercile ranked by the relative basis outperform those in the bottom tercile by 76 bps (t-statistic = 3.83) in the next month, with an annualized Sharpe ratio of 0.61. Controlling for common factors in the commodity futures market, including the traditional basis factor, has little impact on our result; the monthly return spread remains economically important and statistically significant at 59 bps (t-statistic = 3.04). For comparison, the return spread sorted by the traditional basis, after controlling for the relative basis factor and other common factors in commodity futures markets, is a statistically insignificant 6 bps (t-statistic = 0.31) per month.

We then conduct additional tests to further establish the link between our finding about the relative basis return predictability and the Theory of Storage. First, we examine the return prediction power of positive and negative relative basis separately. The Theory of Storage suggests that the relationship between inventory and convenience yield should be convex, that is, the convenience yield increases at an increasing rate when inventory decreases and approaches the stock-out level. Therefore, the convenience yield, which can be interpreted as the additional benefit of holding the physical inventory instead of futures positions, should be greater in low-inventory scenario (i.e., positive relative basis) than in high-inventory scenario (i.e., negative relative basis). This is indeed what we find in the data — the coefficient estimate of positive relative basis in the futures return prediction regression is much larger and more significant than that of negative relative basis. Next, we compare the return prediction power of relative basis in the economic expansion and contraction. We hypothesize that the return predictability of relative basis should be stronger in expansion periods. This is because when the economy is in expansion, manufacturers are more likely to receive a large number of orders from their clients and hence have a stronger incentive to keep their production line uninterrupted, by owing sufficient commodity input inventory. It suggests that the convenience yield should be greater in expansion, and so is the return predictability of relative basis. We find that, for the long-and-short portfolio sorted on relative basis, both its raw return and risk-adjusted alpha are substantially higher in the economic expansion period. This observation is consistent with our hypothesis described above.

In a placebo test, we extend our analyses to financial futures contracts (e.g., stock index futures, currency futures, and interest rate futures), which are not subject to physical inventory constraints since many financial futures are settled in numbers and, if needed, investors can create additional supply by short-selling the underlying securities. We find that the relative basis does not predict returns of financial futures contracts. In contrast, consistent with prior studies, the traditional basis is an important and significant predictor of financial futures returns. This result does not come as a surprise: as we argue throughout the paper, the relative basis reflects the demand and supply of the underlying commodity and the level of its physical inventory (or its scarcity) — an argument irrelevant for financial futures. A natural takeaway from this exercise is that unlike the relative basis, the return predictability of the traditional basis is not due to its relation to the level (or scarcity) of inventories, but rather because of some slow-moving asset-specific characteristics, such as the dividend-price ratio and interest rate differential.

Our study contributes to the long-standing literature on the Theory of Storage in commodity futures markets, which dates back to at least Kaldor (1939) and Working (1949).[3] One of the central predictions of the Theory of Storage is that the convenience yield, which is a result of the scarcity of the spot commodity's inventory, accrues to

[3] Later studies about the theory of storage include Brenann (1958), Fama and French (1988), Hirshleifer (1990), Ng and Pirrong (1994), Carbonez, et al. (2009), Gorton, Hayashi, and Rouwenhorst (2013), etc.

physical commodity owners but not the owners of commodity futures contracts. In the extant literature, the inventory level and associated convenience yield are often measured by the traditional basis, which is confounded by the cost of carry. In this paper, we propose a more precise measure of the convenience yield – the "relative basis" by taking the difference between the traditional basis and a longer-term basis. We show that the relative basis is more closely related to the temporary variation in commodity inventories, particularly its scarcity, as suggested by Brennan (1958), Deaton and Laroque (1992, 1996), and Routledge, Seppi, and Spatt (2000).

We also contribute to the literature that examines the determinants of risk premia in commodity futures markets.[4] We show that our relative basis measure is a stronger futures return predictor than the traditional basis: in a horse race regression in which we include the relative basis, traditional basis, and a host of known predictors of commodity futures returns, the relative basis subsumes traditional basis in predicting subsequent futures returns. We then show that this finding holds in a variety of robustness checks; for example, in a series of spanning tests, the relative basis factor (which goes long the top tercile of commodity futures sorted by the relative basis and short the bottom tercile) is more mean-variance efficient than the traditional basis factor (constructed in a similar fashion). Our empirical finding highlights the importance of convenience yield in determining commodity futures returns. It enriches the implication of the Theory of Storage on the pricing mechanism of commodity markets.

The remainder of the paper is organized as follows. Section 2 describes the sample data and empirical method. Section 3 presents the return predictability of the relative basis. Section 4 conducts portfolio sorting analysis based on the relative basis. Section 5 provides further detailed analysis and Section 6 concludes.

[4] More particularly, this literature includes the following studies: Fama and French (1987), Rouwenhorst and Tang (2012), Gorton, Hayashi, and Rouwenhorst (2013), and Szymanowska, et al. (2014) who document that higher basis leads to higher commodity returns; Bessembinder (1992) and Kang, Rouwenhorst, and Tang (2020) show that hedging pressure has a strong predictive power for futures returns, which is consistent with the theory of normal backwardation; Erb and Harvey (2006) and Miffre and Rallis (2007) find a momentum effect in commodity markets; Boons and Prado (2019) document that a basis momentum measure can forecast commodity futures returns.

2. Data and Methodology

We obtain monthly data on futures prices from Commodity Systems Inc. for all commodities traded on NYMEX, CBOT, and CME. Our sample covers 24 commodities for the period from January 1979 to June 2018. We compute the excess return of commodity i in month t using the price of the nearest futures contract that does not expire in month t:

$$ret_{i,t} = \frac{F_{i,t}(T) - F_{i,t-1}(T)}{F_{i,t-1}(T)}, \tag{1}$$

where $F_{i,t}(T)$ is the futures price of commodity i at the end of month t for the nearest contract with expiration date T and $F_{i,t-1}(T)$ is the price of the same contract at the end of month t-1.

For the basis measure, we follow the related literature and define the Traditional Basis (*TradtBasis*) as the log-difference in prices between the nearest and second-nearest futures contracts, scaled by their difference in time to maturity (annualized), as below:

$$TradtBasis_{i,t} = \frac{ln(F_{i,t}(T1)) - ln(F_{i,t}(T2))}{T2 - T1} \tag{2}$$

where $F_{i,t}(T1)$ and $F_{i,t}(T2)$ are futures prices for commodity i at the end of month t for futures contracts with expiration dates $T1$ and $T2$, respectively.

Next, we propose a new basis measure, dubbed the Relative Basis (*RelatBasis*), and which is defined as the difference between the traditional basis measure introduced above and a longer-term basis measure based on the futures contracts with longer expiration dates. More specifically, the relative basis is defined as:

$$RelatBasis_{i,t} = \frac{ln(F_{i,t}(T1)) - ln(F_{i,t}(T2))}{T2 - T1} - \frac{ln(F_{i,t}(T2)) - ln(F_{i,t}(T3))}{T3 - T2} \tag{3}$$

where $F_{i,t}(T1)$, $F_{i,t}(T2)$, and $F_{i,t}(T3)$ are futures prices at the end of month t for the first nearby, second nearby, and third nearby futures contracts with expiration dates $T1$, $T2$, and $T3$, respectively.[5]

We then conduct a cross-sectional regression of the traditional basis on the relative basis in each month:

$$TradtBasis_{i,t} = a_t + b_t * RelatBasis_{i,t} + \varepsilon_{i,t}\ . \tag{4}$$

From the regression above, we define the Residual Basis (*ResidBasis*) as the sum of the intercept and residual terms of equation (4). Therefore, the relative basis and residual basis are the two independent components that will add up to the traditional basis by construction.

We also construct various commodity futures characteristics that are known to forecast commodity futures returns. These characteristics include the price momentum s (Miffre and Rallis 2007; Asness, Moskowitz, and Pedersen 2013), the basis momentum (Boons and Prado, 2019), and the smoothed hedging pressure (Kang, Rouwenhorst, and Tang, 2020). The price momentum measure (*Momentum*) is calculated as the commodity's past twelve-month cumulative returns. The basis momentum (*BasisMom*) is the difference between the price momentum measures of the first-nearby and second-nearby futures contracts. The smoothed hedging pressure (*SHP*) is defined as the average of the net short position (short position minus long position) of the commercial traders in the Commitments of Traders (COT) report from CFTC in the past one year, scaled by the commodity's most recent open interest.[6] Details of the construction of all these control variables are included in Appendix Table A1.

[5] In computing the relative basis measure, all three contracts (most-nearby, second-nearby, and third-nearby) are assumed to have good liquidity, and hence the prices of these three contracts reflect fresh market information. As is commonly known, the nearby and second-nearby futures contracts have good liquidity; it is hence necessary to check the liquidity of the third-nearby contract. As reported in Appendix Table A2, the volume and open interest ratio of the third-nearby contract to the second one are approximately 30% and 40%, respectively, which indicates active trading in the third-nearby contract. Note that the first and second contracts have similar magnitudes of trading volume and open interest.

[6] We obtain the position measures for various types of participants in commodity futures markets by using the publicly available Commitment of Trader (COT) data provided by the commodity futures trading commission (CFTC). The COT data report the aggregate long and short positions of various

We also follow Gorton, Hayashi, and Rouwenhorst (2013) and collect monthly inventory data from a variety of sources for our sample of commodities.[7] Since there could be a time trend in the inventory data for many of our commodities, we define the normalized inventory level ($Inventory_{i,t}$) as the ratio of the actual inventory to its past one-year average:

$$Inventory_{i,t} = \frac{RawInventory_{i,t}}{\sum_{k=t-12}^{t-1} RawInventory_{i,k}/12} \tag{5}$$

where $RawInventory_{i,t}$ is the actual inventory level of commodity i in month t. Our monthly inventory data cover the period from January 1993 to June 2018. The mean of the normalized inventory level is 1.023, with a standard deviation of 35.0%, indicating large fluctuations in inventories over time.

In Table 1, Panel A reports the time-series averages and standard deviations of the relative basis, traditional basis, and monthly futures returns. Our sample commodities are sorted into five categories: energies, metals, softs, grains, and live stocks. Panels B presents the cross-sectional correlations between the relative basis and other commodity characteristics. For example, the relative basis has a correlation of 0.51 with the traditional basis, which suggests that although these two basis measures are related to each other, they are *not* the same. The relative basis has a low correlation of 0.21 with the basis momentum and virtually zero correlation with the price momentum and the smoothed hedging pressure measures.[8] These correlation coefficients suggest that our novel relative basis measure contains independent information with respect to other well-known commodity characteristics. In Panel C, we report the autocorrelations of the three basis variables. Our relative basis measure has an autocorrelation of 0.33 with one lag, and its autocorrelation becomes statistically insignificant with more lags. In contrast, the traditional basis is much more persistent: its first-order autocorrelation is

---

types of commodity futures market participants, that is, commercial traders, non-commercial traders, and non-reportables. These positions are measured every week on Tuesday and are publicly released three days later, after the market closes on Friday of the same week. We aggregate the weekly COT traders' position data at the monthly level to match our return and basis measures.

[7] The detailed description of these inventory data is available in Appendix B of Gorton, Hayashi, and Rouwenhorst (2013).

[8] The relative basis has zero correlation with the residual basis by definition.

as high as 0.68, which drops to somewhere between 0.3 and 0.5 with more lags, but remain highly significant. The residual basis is even (slightly) more persistent. Put together, these observations suggest that the relative basis captures the fast-moving component of the traditional basis and hence is a better proxy of the temporary nature of inventory shocks in commodity markets.

3. Return Predictability of the Relative Basis

One important implication of the theory of storage (Working, 1949 and Brennan, 1958) is that since commodities are often inputs for production, immediate ownership of a commodity is preferred compared to futures ownership, especially when the commodity is in short supply. In other words, the owners of physical commodity inventory can receive an implicit but important benefit, namely the convenience yield, which cannot be received by investors who buy commodity futures. The convenience yield can benefit the physical commodity owners via the following channels: (a) it helps to ensure the continuity of production with continuous commodity inputs, which is particularly important if the commodity is likely to stockout. For example, when facing a low commodity supply, manufacturers that use the commodity as an input for their production need to consider how to fill their clients' orders in time and how to avoid the restarting cost of their manufacturing line. In this scenario, keeping sufficient inventory of the physical commodity will bring important benefits that cannot be achieved by taking long positions in commodity futures. (b) A commodity supply decrease typically results in a surge in commodity prices, which may in turn lead to increases in the prices of the goods that are produced based on this commodity. If the manufacturer owns sufficient commodity inventory when the shock arrives, she can take this opportunity to sell her products at a better price and increase the market share for her business if her competitors are not as well-prepared as she is. Apparently, this is also a benefit that is available only to physical commodity owners, but not to futures buyers.

Therefore, since futures buyers cannot receive all these benefits of owning physical commodities (i.e., the convenience yield) as illustrated above, they need to be compensated with a return premium. This return premium should be greater when the inventory level is lower. As shown by Brennan (1958), Brennan (1991), Pindyck (1994), and Gorton, Hayashi, and Rouwenhorst (2013), the convenience yield is a convex decreasing function of the inventory level, as depicted in Figure 1. However, the convenience yield in reality is usually difficult to obtain because the actual inventory data are often difficult to acquire and are only available with severe delays.

Financial economists typically use the traditional basis as a proxy for the inventory level to forecast futures returns. Previous studies, such as Fama and French (1987), Rouwenhorst and Tang (2012), and Szymanowska et al. (2014), show that the traditional basis measure indeed positively forecast commodity futures returns. However, as we have discussed before, the traditional basis reflects not only the convenience yield but also other confounding factors, such as storage cost and financing cost, which can vary substantially across commodities. In this section, we propose a new measure of the convenience yield, the relative basis, and conduct a horserace to compare the cross-sectional return predictability of the relative basis, traditional basis, and residual basis.

3.1 Theoretical Motivation

The classical cost-of-carry formula suggests that the relation between commodity futures and spot prices (see Fama and French, 1988, etc.) is prescribed by the equation below:

$$F_{i,t}(T) = S_{i,t}\exp\left[r_{i,t}(t,T) - \delta_{i,t}(t,T) + w_{i,t}(t,T)\right] \tag{6}$$

where $F_{i,t}(T)$ is the future price for commodity i at time t with maturity T, $S_{i,t}$ is the spot price for commodity i at time t, $r_{i,t}(t,T)$ is the financing cost of commodity i from t to T, $\delta_{i,t}(t,T)$ is the convenience yield of commodity i from t to T, and $w_{i,t}(t,T)$ is the storage cost of commodity i from t to T.

In the existing literature, financial economists use the difference between $S_{i,t}$ and $F_{i,t}(T)$ (or between $F_{i,t}(T1)$ and $F_{i,t}(T2)$), namely the traditional basis measure, as a proxy for the convenience yield as implied by the Theory of Storage, and also use this measure to predict commodity futures returns. However, as shown in equation (6) above, the traditional basis measure includes other factors (i.e., the storage cost and the financing cost) more than the convenience yield itself. For the storage cost, there could be substantial difference in it across different commodities. First, the storage cost can vary by the nature of commodity: imagine the difference between storage costs of gold and natural gas — the storage cost of the latter is larger than the former because of the warehousing cost and environmental reasons. What's more, even for a specific commodity *i*, its $w_{i,t}$ can vary over time because the price of commodity may also change at different time point *t*.[9] For the financing cost, it may vary across different commodities as well, because some commodities (e.g., gold) can be more readily pledged as collateral for borrowing than other commodities (e.g., live cattle). Therefore, all these differences in the storage and financing costs among commodities can substantially undermine the effectiveness of the traditional basis as a proxy for the convenience yield in commodity markets.

To provide a cleaner and a more precise proxy for the convenience yield, we construct the relative basis measure as the difference between two close-by traditional basis measures as follows:

$$RelatBasis_{i,t} = TradtBasis_{i,t}(T1, T2) - TradtBasis_{i,t}(T2, T3) \ . \tag{7}$$

where

$$TradtBasis_{i,t}(T1, T2) = \frac{1}{T2-T1}[\delta_{i,t}(T1, T2) - r_{i,t}(T1, T2) - w_{i,t}(T1, T2)], \tag{8}$$

$$TradtBasis_{i,t}(T2, T3) = \frac{1}{T3-T2}[\delta_{i,t}(T2, T3) - r_{i,t}(T2, T3) - w_{i,t}(T2, T3)]. \tag{9}$$

In equation (7), the "noise" part (i.e., the storage and financing costs) in the traditional basis measure used by the extant literature is canceled out by taking the difference between equations (8) and (9). More specifically, for a certain commodity *i* at time *t*, assuming *T1*, *T2*, and *T3* are not too far away from each other, the two storage

[9] In equation (6), $w_{i,t}$ can be understood as the physical cost for the storage of one unit of the underlying commodity divided by the corresponding commodity price. Therefore, it is influenced by both the difference in physical storage cost and the fluctuation of commodity price.

costs $w_{i,t}(T1,T2)$ and $w_{i,t}(T2,T3)$ should be quite similar to each other.[10] Moreover, the financing cost for a given commodity should be also persistent along time, that is, the financial costs for commodity i from $T1$ to $T2$ and from $T2$ to $T3$ should be very close to each other. Therefore, by taking the difference between equations (8) and (9) above, we essentially remove the interference of storage cost and financing cost on the effectiveness of the basis measure as a proxy for the convenience yield.

Our relative basis measure can capture the "signal" part in the convenience yield that responds to the mean-reversion nature of the commodity inventory. Liu and Tang (2011) show that inventory follows a mean-reverting process – when the inventory in the current period is low, it will lead to an increase in commodity prices that in turn reduces consumption and stimulates production for the commodity, and hence the inventory level is likely to rebound in the next period. Consistent with Liu and Tang (2011), we and other studies find that commodity inventories are indeed mean-reverting at the monthly horizon. Similar to inventories, Schwartz (1997), Casassus and Collin-Dufresne (2005), and Liu and Tang (2011) suggest that the convenience yield should also be mean-reverting. As shown in Figure 1, if at time t, the short-term convenience yield is high because of the shortage of inventory in the near-term future (point A), investors can expect a lower next-period convenience yield, as the inventory will fill up, as explained above (point B). In this scenario, the relative basis measure is positive as the reflection of inventory scarcity. Similarly, when inventory is high and the convenience yield is low, the relative basis should be negative (as illustrated in point C). Therefore, the relative basis should have a positive relationship with the convenience yield and a negative relationship with the inventory level.

Overall, the relative basis measure is a more precise proxy for the convenience yield that stems from the inventory scarcity as suggested by the Theory of Storage. Thus, it should be a more powerful return predictor on commodity futures markets. This is what we are going to test in the next subsection.

[10] This is because, as long as $T1$, $T2$, and $T3$ are not too far away from each other, the expected physical storage costs and the expected commodity prices between these time points (i.e., from $T1$ to $T2$ and $T2$ to $T3$) should be close to each other.

3.2 Relative Basis and Return Predictability in Commodity Futures Markets

We run Fama-MacBeth cross-sectional regression of commodity futures returns in month t+1 on various basis measures (the traditional basis, relative basis, and residual basis), as well as other control variables, in month t as follows:

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}ResidBasis_{i,t}(\text{ or } TradtBasis_{i,t}) + \lambda controls_{i,t} + \varepsilon_{i,t} \quad (10)$$

where $ret_{i,t+1}$ is the futures returns for commodity i in month t+1. Control variables are those considered as asset pricing characteristics (besides basis) in commodity futures markets, including the price momentum (Asness, Moskowitz and Pedersen, 2013; Bakshi, Gao, and Rossi, 2017) and basis momentum (Boons and Prado, 2019). The price momentum control variable is constructed as the cumulative past twelve-month returns, and the basis momentum is defined as the difference between the two price momentums based on the first-nearby and second-nearby futures contracts. We use the Newey-West method to adjust the t-statistics for the time-series average of the cross-sectional regression coefficient estimates to address potential autocorrelation concerns.

In Panel A of Table 2, columns 1 to 3 report the coefficient estimates from univariate regressions in which the next-month commodity futures returns are regressed on the relative basis, residual basis, and traditional basis one by one in each regression. We find that the regression coefficient is 0.018 (t-statistic =3.21) for the relative basis (as reported in column 1), 0.011 (t-statistic =1.60) for the residual basis (as reported in column 2), and 0.013 (t-statistic =2.53) for the traditional basis (as reported in column 3). Our observation here suggests that the relative basis can significantly forecast commodity futures returns. We also find that when other controls are not included, the residual basis possesses some marginal return predictability. Not surprisingly, the return predictability of the traditional basis, which can be thought of as the combination of the relative basis and the residual basis, is somewhere in between.

More importantly, we find that after we include controls of the price momentum and the basis momentum into the regression, the return predictability of the relative basis is *unaffected*. Column 4 indicates that the coefficient estimate of the relative basis is still 0.018, with a significant t-statistic of 2.61. In terms of economic importance, an increase of relative basis from the average minus one standard deviation to the average plus one standard deviation will increase the futures return by 0.64% per month, or 7.8% per annum. Therefore, the return predictability of the relative basis is both statistically significant and economically important, with or without controls.[11]

In sharp contrast, we find that the return predictability of the traditional and residual basis *disappears* when adding the controls into the regression. In fact, with the controls included, the coefficient estimate of the residual basis becomes negative but insignificant, and the coefficient for the traditional basis becomes essentially zero, as reported in columns (5) and (6), respectively. Next, we run a horserace test in which we put either the relative basis and the residual basis, or the relative basis and the traditional basis, together with other controls in the same return-prediction regression. As reported in columns (7) and (8), the relative basis subsumes the other two basis measures in terms of predicting subsequent commodity futures returns. In Panel B, we employ the next-quarter commodity futures returns as the dependent variable and re-run the same regressions as in Panel A. Similar results are obtained.

For robustness, we also run a panel regression of equation (10) with both time and commodity fixed effects. The results in Appendix Table A3 are consistent with those found in the Fama-Macbeth cross-sectional regressions, which informs that the return predictability of relative basis is robust to different regression methodologies.

In brief, in this subsection we find that the relative basis has strong predictability for commodity futures returns; in contrast, the return predictability of the residual basis and the traditional basis does not exist when including control variables. Moreover, the return predictability of the traditional basis and relative basis is subsumed by the relative basis.

[11] For the control variables, we find that their coefficient estimates are consistent with previous studies, that is, high price momentum and basis momentum predict higher subsequent futures returns.

3.3 Any Return Predictability of Relative Basis in Financial Futures Markets?

In this subsection, we examine the return predictability of the three basis measures for financial futures markets. The reason to redo the test in financial futures markets is mainly that there is no inventory concern for financial futures, and therefore the Theory of Storage should not apply to financial futures. In other words, if the return predictability comes from any variable that relates to the Theory of Storage and the associated concept of convenience yield, we should expect that it *cannot* predict financial futures returns.

Moreover, unlike the convenience yield, which is mean-reverting in commodity markets, the dividends issued from T1 to T2 and from T2 to T3 for financial assets (e.g., stocks) should be very close to each other and hence canceled out in the relative basis measure; therefore, the relative basis tends to be very small in magnitude and should not have the return predictability for financial futures. We compute the mean value of relative basis for 19 commonly-used financial futures in Appendix Table A4. The magnitudes of the relative basis for these financial futures are much lower than those of commodity futures.

Table 3 indicates that in financial futures markets, regardless of whether we include the controls, the relative basis has insignificant coefficient estimates in the return-predicting regressions (see columns 1 and 4). As a comparison, both the residual basis and the traditional basis exhibit strong return predictability in financial futures markets, with or without controls (see columns 2, 3, 5, and 6), which is consistent with Koijen, Moskowitz, Pedersen, and Vrugt (2018). Moreover, when we put the relative basis and the traditional basis (or the residual basis) together, this time it is the traditional basis (or the residual basis) that dominates the relative basis as the predictor for financial futures returns (see columns 7 and 8). In brief, our finding in this section suggests that the return predictability of the relative basis does come from the Theory of Storage, and therefore, should only apply to the commodity markets, but not to financial futures markets.

3.4 Return Predictability of the Standardized Relative Basis

We notice that the size of the relative basis is quite different for commodity futures and financial futures markets. Therefore, to conduct a fair comparison of their economic impact on subsequent futures returns, this subsection employs the standardized version of these basis measures to further compare their return predictability. More specifically, for each month, we standardize the basis measures and control variables by subtracting their cross-sectional mean and then dividing by their cross-sectional standard deviation.

Table 4 employs the Fama-Macbeth regression as follows:

$$ret_{i,t+1}(ret_{i,q+1}) = a_t + b_{1,t}RelatBasis_{i,t}^{adj}$$
$$+b_{2,t}ResidBasis_{i,t}^{adj}(or\ TradtBasis_{i,t}^{adj}) + \lambda_t Controls_{i,t}^{adj} + \varepsilon_{i,t} \quad (11)$$

where $ret_{i,t+1}$ is the cumulative futures returns in month $t + 1$ and $ret_{i,q+1}$ is the cumulative futures returns in quarter $q + 1$. The main explanatory variables are the standardized relative basis, residual basis, and traditional basis, denoted as $RelatBasis^{adj}$, $ResidBasis^{adj}$, and $TradtBasis^{adj}$. The control variables include the standardized price momentum ($momentum^{adj}$) and the standardized basis momentum ($BasisMom^{adj}$).

The left three columns of Panels A and B focus on the commodity markets. Columns (2) and (3) of these two panels indicate that the relative basis subsumes the other two basis measures in terms of predicting subsequent commodity futures returns. When we turn to the latter three columns of two panels (Columns (4) to (6)), the residual basis and the traditional basis dominate the relative basis as the return predictor for financial futures markets. In terms of the economic magnitude, for commodity futures, a one standard deviation increase in $RelatBasis^{adj}$ will cause the return increase by 44 basis points in the next month (column 3 of Panel A) and 99 basis points in the next quarter (column 3 of Panel B). In comparison, for financial futures, $RelatBasis^{adj}$ has little impact on the subsequent returns.

3.5 Robustness Tests

In this subsection, we introduce a set of additional tests to ensure the robustness of our main results. First, we include an additional control variable, the smoothed hedging pressure (*SHP*), into the futures return prediction regression. Following Kang, Rouwenhorst, and Tang (2020), we first obtain the hedging pressure as the net short position of the COT commercial traders and then compute the smoothed hedging pressure as the past one year moving average of the hedging pressure. The inclusion of the smoothed hedging pressure allows us to control for the hedging effect documented in Kang, Rouwenhorst, and Tang (2020) under the classical Keynes' theory of normal backwardation.[12] We use the smoothed hedging pressure in the week at the end of a specific month as the monthly *SHP* measure. Since CFTC started to announce the weekly COT reports from 1993, we consider the sample period that starts from the year 1993 (instead of 1979) to include the *SHP* in the list of control variables.

In the first column of Appendix Table A5 Panel A, we run the regression of next-month commodity futures returns on the relative basis and the control variables, including the price momentum, basis momentum, and smoothed hedging pressure. We find that adding the additional *SHP* control does not alter the predictability of the relative basis. This finding is also robust to the further inclusion of either the residual basis or traditional basis in the regression, as indicated by columns 2 and 3. We also observe that the residual basis or traditional basis still cannot predict commodity futures returns in such a horserace test, which is consistent with our findings in the previous Section 3.2.

In Panel B, we redo the same test for financial futures and find that neither the relative basis nor the smoothed hedging pressure can forecast the return of financial futures. This observation suggests that both the Theory of Storage (as measured by the relative basis) and the Theory of Normal Backwardation (as measured by the smoothed hedging pressure) are unique to commodity markets. Similar to what we find in Section 3.3, the residual basis and the traditional basis both have significant coefficient estimates in predicting financial futures returns. This, again, suggests that the return

[12] Kang, Rouwenhorst and Tang (2020) also propose a liquidity factor, Q, which has predictability within 2 weeks. However, as our paper is mainly on a monthly horizon, we do not include this factor as a control variable.

predictability of these two measures should be due to a reason that is different from the Theory of Storage.

In further robustness tests, we split the full sample period into two subperiods (1979-1999 and 2000-2018) and show that the relative basis measure can predict commodity futures returns in both subperiods, as indicated in Appendix Table A6.

In commodity markets, the length of the difference in maturity between two futures contracts may vary across different commodities. In our sample, most commodities have futures contracts matured every two months, except that the energy section (including crude oil, heating oil, and natural gas) has futures contracts in every month of the year, and two precious metals (platinum and palladium) have the contracts every three months. Thus, in Appendix Table A7, we adjust the maturities of the first, second, and third nearby contract of crude oil, heating oil, and natural gas, so that the relative basis is constructed on the same time interval (two months). [13] Table A7 shows that the return predictive power of relative basis holds when we use this alternative construction method of relative basis.

Combining the findings in this subsection, we can obtain a clearer picture regarding which basis measure is a more precise proxy for the commodity futures risk premia as suggested by the Theory of Storage. In commodity markets, the relative basis is a strong and robust predictor for futures returns, consistent with the prediction of the Theory of Storage. In financial futures markets, since there is little inventory concern and the convenience yield should approach zero, the relative basis should not be able to predict subsequent futures returns. This is indeed what we observe in the data. In contrast, the return predictability of the traditional basis and residual basis is weak (or nonexistent) in commodity futures markets where the Theory of Storage applies, and becomes much stronger for financial futures markets where the Theory of Storage does not apply. Therefore, we conclude that the relative basis is a more precise measure for

[13] Note that platinum and palladium have three-month maturity interval, and contracts beyond 4th maturity are very illiquid; we thus still keep the platinum and palladium contracts unchanged.

the convenience yield implied by the Theory of Storage and has an important asset pricing implication on commodity futures markets.

4. Relative Basis Portfolios

4.1 Portfolio Sorting Analysis

In addition to the Fama-MacBeth regression, we also conduct a portfolio sorting analysis based on these three basis variables as a nonparametric test of our previous main findings. At the end of each month, we construct three equally weighted commodity futures portfolios based on the relative basis. We then observe the returns of these three portfolios, in addition to the return difference (P3-P1) between the portfolios with the highest and lowest relative basis, in the post one month and one-quarter horizons. We also redo the portfolio sorting analysis based on the traditional basis and the residual basis accordingly.

In Panel A of Table 5, we report the portfolio sorting results for commodity futures markets. We find that commodities with high relative basis deliver significantly higher returns than those with low relative basis. For example, in the post one-month horizon, the holding returns of portfolios P1 (low relative basis) and P3 (high relative basis) are −0.21% and 0.56%, respectively. The long-short strategy of buying P3 and selling P1 produces an average return of 0.76% per month, with a *t*-statistic of 3.83 and an annualized Sharpe ratio of 0.61. Similarly, during the post one-quarter horizon, the holding returns of portfolios P1 and P3 are −0.93% and 1.87%, respectively. The strategy of taking long position in P3 and short position in P1 yields an average return of 2.79% per quarter, with a *t*-statistic of 6.83 and an annualized Sharpe ratio of 0.72. We also find a significant return difference between commodity futures with high and low residual (or traditional) basis. This observation is consistent with the previous univariate regression results.

In Panel B of Table 5, we redo the portfolio sorting analysis for financial futures. We find that sorting on the relative basis yields little difference in return between high and low relative-basis portfolios, while the return difference between high and low traditional- or residual-basis portfolios is significantly positive. This is consistent with what we observe in Section 3.3 in terms of the regression test in financial futures markets.

In brief, we find that the raw return differences between high and low basis portfolios are significant for all three basis measures on commodity futures markets. A probably more important question is whether their risk-adjusted alpha can remain significant. This is what we are going to examine in the next subsection.

4.2 Asset Pricing Test for Relative Basis Portfolio

In this subsection, we first obtain the time-series returns of the high-minus-low portfolio for the relative basis, the residual basis, and the traditional basis, as introduced in the previous section. We then construct factor portfolios based on the price momentum and the basis momentum in a similar manner (i.e., the highest tercile minus the lowest tercile) and obtain the time-series returns of these factors as controls. We also construct the market factor of equally weighting returns of 24 sample commodity futures. As shown in Appendix Table A8, the correlations between the relative-basis portfolio return and other factor returns are low. This result suggests that the return difference between high and low relative-basis portfolios is unlikely to be explained by other factors.

We run an asset pricing test by regressing the relative-basis portfolio returns on the contemporaneous market, price-momentum, and basis-momentum factor returns, as reported in Table 6. The first three regressions in Table 6 indicate that the relative-basis portfolio return is statistically significant (at the 1% level) and economically important (approximately 0.6% per month), even after controlling for other factors and including the traditional basis or residual basis factors returns. The magnitude of the relative-basis factor alpha here also coincides well with the economic magnitude gauge conducted in Section 3.2, where we show that an increase in the relative basis from the average minus one standard deviation to the average plus one standard deviation will lead to an increase in the next-month return by 0.64%.

In contrast, the returns of the factor portfolios based on the residual or traditional basis become small and insignificant after including the control factors, i.e., the market, price momentum, basis momentum, and the relative basis factors. For example, the alpha of the traditional basis factor portfolio is 0.26%, with a t-statistic of 1.30, when other controlling factors except the relative basis are included. It further decreases to only 0.06%, with a t-statistic of merely 0.31, when the relative basis factor is added. The residual-basis factor portfolio has an insignificant alpha of 0.11% in both regression specifications. Therefore, our observation here suggests that for the traditional basis and residual basis, although their factor portfolio has a significant raw return, the return can be almost fully explained by other already-known factors and the relative basis factor. This observation is consistent with the characteristic-based return-prediction regression results presented in Section 3.2.

4.3 Portfolio Returns and Business Cycle

Section 3 discusses that the convenience yield can benefit the physical commodity owners via two channels. First, it helps to ensure the continuity of the production with uninterrupted commodity inputs, and this is particularly important if the commodity is more likely to stockout. Second, a commodity supply decrease typically results in a surge of commodity prices, which may in turn lead to increases in the prices of the goods that are produced based on this commodity. Since futures buyers cannot receive all these benefits of owning physical commodities (i.e., the convenience yield), they need to be compensated by earning a return premium by holding futures positions instead of spot positions. We argue that such return premia should be greater during economic expansion periods. This is because, in expansion, manufacturers are more likely to receive a large amount of orders from their clients and hence have a stronger incentive to keep their orders filled in time with uninterrupted commodity inputs and production plans. It suggests that the convenience yield should be greater in economic expansion, and so is the return predictability of relative basis.

Table 7 examines the return predictability of relative basis among different business cycles. The business cycle index is proxied by the Chicago Fed National Activity Index (CFNAI).[14] The full sample is equally split into two subsamples: contraction and expansion periods. Panel A conducts the portfolio sorting based on the relative basis and reports the difference in return spreads of low and high relative-basis groups between two periods. Panel B reports the alpha of relative-basis portfolio returns controlling for other major factors for expansion and contraction periods separately. Both panels indicate that the return predictability of the relative basis is substantially greater in expansion periods. For example, as shown in Panel A, the strategy of buying P3 and selling P1 produces an average return of 1.09% per month in expansion periods and 0.44% per month in contraction periods. Even after controlling for other factors (in Panel B), the alphas of the relative-basis portfolio are 0.97%(t-statistic=3.34) and 0.33%(t-statistic=1.17) in the expansion and contraction periods, respectively. The results in Table 7 are consistent with our argument that the return premia based on the relative basis should be larger in economic expansion.

5. Further Analysis

5.1 Asymmetric Return Predictability of the Relative Basis

As shown in Figure 1, the Theory of Storage (e.g., Brennan, 1958, Pindyck 2001, and Deaton and Laroque, 1992) proposes a convex relationship between inventory and convenience yields. Specifically, at low inventory levels, the convenience yield increases at an increasing rate as the inventory depletes. When the inventory is close to a stockout, the additional benefit of holding the spot inventory instead of futures positions becomes much greater than in the high-inventory-level case. Thus, futures prices will be much more discounted relative to spot prices, and investors who hold long positions in futures should take a higher return premium than that in a high-inventory scenario.

[14] Refer to https://www.chicagofed.org/research/data/cfnai/current-data

We hence hypothesize that positive relative basis, which represents the lack of inventory, should have more prominent futures return predictive power than negative relative basis. To test this relationship, we run the following Fama-MacBeth regression:

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis\_pos_{i,t} + b_{2,t}RelatBasis\_neg_{i,t} + b_{3,t}ResidBasis_{i,t}(or\ TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t} \quad (12)$$

where we set $RelatBasis\_pos_{i,t}$($RelatBasis\_neg_{i,t}$)as the relative basis of commodity i in month t if it is positive (negative), and zero otherwise. The control variables in this regression include the price momentum and basis momentum, as in previous regressions.

Table 8 indicates that in all regression specifications, the coefficient estimates of $RelatBasis\_pos_{i,t}$ are significantly positive, whereas those of $RelatBasis\_neg_{i,t}$ are positive but insignificant and much smaller in terms of magnitude. Therefore, our observation suggests that it is the positive relative basis, which is associated with the depletion of inventory and increasing convenience yield arose from the inventory stockout, that drives the return predictability of relative basis.

5.2 Samples Partitioned on the Absolute Value of the Relative Basis

We hypothesize that the return predictability of relative basis is related to its magnitude (i.e., the absolute value of relative basis). The low magnitude of relative basis suggests that commodity is less likely to experience the commodity stockout and hence its convenience yield should be less important. So we expect that the relative basis should have more(less) return predictive power in commodity with high (low) absolute value of relative basis.

Table 9 examines the return predictability of relative basis among two subsamples partitioned on the absolute value of the relative basis, *abs(RelatBasis)*. The full sample is equally split into two subsamples based on *abs(RelatBasis)* at the end of month *t*: high and low *abs(RelatBasis)* subsamples. We conduct the baseline regression within each subsample. The dependent variables are next month (or next quarter) commodity futures returns. The explanatory variables are the relative basis and the traditional basis (the residual basis). Control variables include price momentum and basis momentum. To conduct fair compassion, for each month, we standardize the variables by subtracting their cross-sectional mean and then dividing by their cross-sectional standard deviation. The regression result suggests that the return predictability of relative basis is concentrated in the commodities with high absolute value of relative basis. This observation is consistent with our hypothesis described above.

5.3 Predicting Spread Returns

In this subsection, we conjecture that the relative basis contains information about futures spread premia. If at time *t*, the short-term convenience yield is high because of the shortage of inventory in the near-term future (point A in Figure 1), investors can expect a lower next-period convenience yield, as the inventory will fill up, as explained above (point B in Figure 1). Over this time, $RelatBasis_t$ is positive, and the risk premium is high due to the inventory shortage. Meanwhile, since the spot convenience yield is greater than the next-period one, the risk premium earned by the nearby futures should also be greater than that of the next-maturity futures, and hence the nearby and next-maturity futures spread returns should be positive. Similarly, if the short-term convenience yield is low due to a high level of inventory, point C in Figure 1, the expected next-period convenience yield is greater than the current convenience yield due to the expected lower production and greater consumption (point B in Figure 1) in the future; therefore, the $RelatBasis_t$ is negative, and the risk premium is small because of the large inventory. Following the same token, the futures spread returns should be negative.

Since the spread trading strategy is also an important trading strategy on commodity futures markets, we empirically examine the futures return spread predictability based on the relative basis with the following regression:

$$ret_{spread_{i,t+1}} = a_t + b_{1,t} RelatBasis_{i,t}$$
$$+ b_{2,t} ResidBasis_{i,t} \left(or\ TradtBasis_{i,t} \right) + \lambda_t controls_{i,t} + \varepsilon_{i,t} \tag{13}$$

where $ret_{spread_{i,t+1}}$ is the difference in futures returns between the first-nearby and second-nearby contracts for commodity i in month t+1. All basis measures and control variables are defined in the same way in the previous section. As shown in Table 10, among all regression specifications, the coefficient estimates of relative basis are significantly positive. This implies that a high relative basis leads to high spread returns on commodity futures markets, which is consistent with our argument above. As a comparison, the traditional basis and the residual basis have almost zero predictability for spread returns. This observation again highlights the importance of the relative basis in pricing the futures contracts on commodity markets.

5.4 Relative Basis and Inventories

To better understand the relationship between various convenience-yield proxies and inventories, we conduct the following Fama-MacBeth regression, in which we include both the inventory itself and the scarcity of inventory as the explanatory variables for various convenience-yield proxies:

$$RelatBasis_{i,t} = a_t + b_{1,t}\, Inventory_{i,t} + b_{2,t}(1/Inventory_{i,t}) + \varepsilon_{i,t} \qquad (14)$$

where $Inventory_{i,t}$ is defined in Equation (5). Following Brennan (1958) and Geman and Nguyen (2005), we use the inverse of inventory, $1/Inventory_{i,t}$, as the proxy of the scarcity for a certain commodity. The rationale is that when inventory becomes low or even approaches zero (i.e., close to stockout), this scarcity measure will increase more than proportionally to reflect the nonnegative bond of the physical commodity inventory. On the left-hand side of the regression, in addition to the relative basis, we also consider the traditional basis and the residual basis.

Panel A of Table 11 indicates that the relative basis is more closely related to the scarcity than the inventory level: the regression coefficients are 0.030 (t-statistic =1.28) for *Inventory* and 0.097 (t-statistic =3.98) for *1/Inventory*, respectively. This implies an upsurge in the relative basis when close to a stockout. Meanwhile, the traditional basis and residual basis mainly relate to the inventory level instead of scarcity. For example, the regression coefficient for the traditional basis is -0.073 (t-statistic =-2.60) for *Inventory* and 0.058 (t-statistic =2.08) for *1/Inventory*, and for the residual basis, it is -0.071 (t-statistic =-3.70) for *Inventory*, and it is approximately zero for *1/Inventory*.

We also examine how different basis measures can predict subsequent inventory changes. As reported in the Appendix Table A9 Panel C, when the relative basis increases, the inventories tend to increase in the next one to three months. This is consistent with our previous argument that a high relative basis is related to low current inventory, and because of the mean-reverting feature (Liu and Tang 2011), the inventory in the next several months tends to rebound (or to be replenished) from current low levels. In contrast, a high traditional basis or residual basis predicts a lower inventory in the next period. This might be explained by the evidence of Bailey and Chan (1993), that inventories have a persistent component that relates to macroeconomic variables or business cycles. In summary, the relative basis is more likely to relate the short-term depletion (stockout) and replenishment of inventories than the traditional basis and residual basis.

Next, we decompose the relative basis into two components: one component that can be explained by the available inventory information and another component that cannot be explained by inventory information. More specifically, we employ a regression of the relative basis on inventory and its inverse value, as described in equation (14), to decompose the relative basis into the inventory-related component ($RelatBasis_{INV}$), the fitted value of the regression, and the inventory-unrelated component ($RelatBasis_{NOINV}$), the residual from the regression.

We examine whether the return predictability of the relative basis comes from its inventory-related or -unrelated component by replacing the relative basis with these two components in the return predictive regression:

$$
\begin{aligned}
ret_{i,t+1} = a_t + b_{1,t} RelatBasis_{INV\,i,t} + b_{2,t} RelatBasis_{NOINV\,i,t} \\
+ b_{3,t} ResidBasis_{i,t} (or\ TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t} \quad (15)
\end{aligned}
$$

Panel B of Table 11 indicates that the return-predictability of the relative basis mainly comes from its inventory-related component. The coefficient estimate of $RelatBasis_{INV}$ is significantly positive in all regression specifications, while the inventory-unrelated component has a much smaller and insignificant coefficient estimate. This suggests that the return-predicting capability of the relative basis does come from the inventory-related information contained in this measure, which is consistent with the Theory of Storage.

Overall, our observation here suggests that the relative basis of a commodity can be viewed as a real-time reflection of the inventory information for this commodity. As such, the relative basis can provide valuable information to market participants, as well as policymakers who may have concerns about the shortage of various commodities that are important to the well-functioning of the economy and the society. This is especially relevant in light of the difficulties in the measurement of the actual demand and supply (or inventory levels) of most commodities -- it is usually done with significant delays and based on noisy, indirect, and incomplete data (see Gorton, Hayashi, and Rouwenhorst, 2013).[15] This result highlights the function of financial markets as the venue to aggregate information from different investors and makes the market price (or price-based proxy) a crystal ball for the underlying fundamentals.

5.5 Robustness Checks of Storage Cost

When constructing the relative basis measure, we assume that the storage cost is fairly persistent within the first to third nearby futures tenures. However, we understand that there exists the possibility that storage cost may increase substantially during the period of exhaustion of storage capacity. Therefore, we check for the robustness of our baseline regression during extreme high inventory periods, by conducting Fama-Macbeth return-predicting regression as follows:

[15] More specifically, researchers usually have to wait for at least a few months in terms of time lag to be able to collect the commodity inventory data from various sources. Some commodity inventory data are not updated through the most recent years as well. For example, the wheat inventory data are only updated through the end of 2011. Moreover, inventories from tankers on the high sea or from developing countries are not possible to collect.

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}Dummy\_HighInv_{i,t}$$
$$+b_{3,t}RelatBasis_{i,t} * Dummy\_HighInv_{i,t} + \lambda_t controls_{i,t} + \varepsilon_{i,t} \qquad (16)$$

where $Dummy\_HighInv_{i,t}$ is set to be one if the inventory of commodity i in month t exceeds the 90th percentile of its time-series distribution, otherwise it is zero. The interaction item in equation (16) examines whether there is any additional return predictability of relative basis when the storage cost might be high. As shown in Appendix Table A10, the coefficient estimate of this interaction item is fairly small and insignificant, while the general return predictability of relative basis is still robust. This observation suggests that high inventory does not disturb the general return prediction power of relative basis.

6. Conclusion

In this paper, we propose a novel measure, dubbed relative basis, as a more precise proxy for the convenience yield implied by the Theory of Storage. Compared with the traditional basis, the relative basis purges out the influence from the storage cost and financing cost, by taking the difference between the short-term and longer-term traditional basis. Consistent with the Theory of Storage, we find that the relative basis is more closely related to the scarcity of inventories than the traditional basis, and it also better captures temporary fluctuations in commodity inventories.

We then examine the return predictability of the relative basis on commodity futures markets. In our Fama-MacBeth regression analysis, we find that although the traditional basis can predict commodity futures returns in univariate regressions, it loses its predictive power when other known return predictors are included. In contrast, the return predictability of the relative basis remains intact with these additional controls. In our portfolio sorting approach, we find that commodities with high relative basis outperform those with low relative basis by 0.6% per month, which cannot be explained by the commodity market, price momentum, basis momentum, and traditional basis factors. Moreover, it is the positive relative basis that drives the return predictive pattern, consistent with the nonnegativity constraint on inventories.

Another interesting result is that the relative basis does not forecast returns of financial futures, but the traditional basis does. Given that there is inventory concern for financial assets, this result further confirms that the return predictability of our relative basis measure does stem from the Theory of Storage, which uniquely applies to commodity futures markets.

In sum, our paper shows that the relative basis is an important and priced characteristic that captures the convenience yield originating from the scarcity of spot inventories and the associated risk premium in commodity futures markets. In a recent paper, Koijen et al. (2018) show that carry (similar to the traditional basis) is an important priced factor across multiple asset classes. Our results, in particular the contrast between commodity futures and financial futures, suggest that the return predictability of carry can arise for a variety of different reasons in different markets.

References

Acharya, Viral V., Lars A. Lochstoer, Tarun Ramadorai, 2013, Limits to arbitrage and hedging: Evidence from commodity markets, *Journal of Financial Economics* 109, 441-465.

Asness, Clifford, Tobias Moskowitz, and Lasse Pedersen, 2013, Value and Momentum Everywhere, *Journal of Finance* 68, 929-985.

Bailey, Warren, K.C. Chan,1993, Macroeconomic Influences and the Variability of the Commodity Futures Basis, *Journal of Finance* 48, 555-573.

Bakshi, Gurdip, Xiaohui Gao and Alberto G. Rossi, 2017, Understanding the Sources of Risk Underlying the Cross Section of Commodity Returns, *Management Science* 65, 1–23.

Bessembinder, Hank, 1992, Systematic risk, hedging pressure, and risk premiums in futures markets, *Review of Financial Studies* 4, 637–667.

Boons, Martijn and Melissa Porras Prado, 2019, Basis-Momentum, *Journal of Finance* 74, 239-279.

Brennan, Michael, 1958, The supply of storage, *American Economics Review* 48, 50–72.

Brennan, Michael, 1991, The price of convenience and the valuation of commodity contingent claims, in: D. Lund and B. Oksendal (eds), Stochastic Models and Option Values, Emerald Group Publishing, North Holland (Amsterdam).

Casassus, Jaime and Pierre Collin-Dufresne, 2005, Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates, *Journal of Finance* 60, 2283-2331.

Deaton, Angus and Guy Laroque, 1992, On the Behavior of Commodity Prices, *Review of Economic Studies* 59, 1-23.

Deaton, Angus and Guy Laroque, 1996, Competitive Storage and Commodity Price Dynamics, *Journal of Political Economy* 104, 896-923.

De Roon Frans, Theo E. Nijman, and Chris Veld, 2000, Hedging pressure effects in futures markets, *Journal of Finance* 55, 1437–1456.

Erb, Claude and Campbell Harvey, 2006, The Strategic and Tactical Value of Commodity Futures, *Financial Analysts Journal* 62, 69-97.

Eastham, J.K., 1939, Commodity Stocks and Prices, *The Review of Economic Studies*, 6, 100–110.

Fama, Eugene F. and Kenneth R. French, 1987, Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage, *Journal of Business* 60, 55–73.

Fama, Eugene F., and Kenneth R. French, 1988, Business cycles and the behavior of metal prices, *Journal of Finance* 43, 1075–1093.

Geman, Hélyette and Vu-Nhat Nguyen, 2005, Soybean Inventory and Forward Curve Dynamics, *Management Science* 51, 1076–1091.

Gorton, Gary B., and K. Geert Rouwenhorst 2006, Facts and Fantasies about Commodity Returns, *Financial Analysts Journal* 62, 47-68.

Gorton Gary, Fumio Hayashi, and K. Geert Rouwenhorst 2013, The fundamentals of commodity futures returns, *Review of Finance* 17, 35-105.

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

Kang, Wenjin, K Geert Rouwenhorst, and Ke Tang, 2020, A Tale of Two Premiums: The Role of Hedgers and Speculators in Commodity Futures Markets, *Journal of Finance* 75, 377-417.

Koijen, Ralph, Tobias Moskowitz, Lasse H. Pedersen, and Evert Vrugt, 2018, Carry, *Journal of Financial Economics* 127, 197-225.

Liu, Peng and Ke Tang, 2011, The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield, *Journal of Empirical Finance* 18, 211-224.

Miffre, Joelle and Georgios Rallis, 2007, Momentum strategies in commodity futures markets, *Journal of Banking and Finance* 31, 1863-1886.

Newey, Whitney K. and West, Kenneth D.,1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pindyck, R. (1994) Inventories and the short-run dynamics of commodity prices, RAND Journal of Economics 25, 141 – 159.

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

Routledge, B., Seppi, D., and Spatt, C., 2000, Equilibrium forward curves for commodities, *Journal of Finance* 55, 1297–1338.

Rouwenhorst, K. Geert, and Ke Tang, 2012, Commodity investing, *Annual Review of Financial Economics*, 447-467.

Schwartz, Eduardo, 1997, The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging, *Journal of Finance* 52, 923-973.

Szymanowska, Marta, Frans De Roon, Theo Nijman and Rob Van Den Goorbergh, 2014 An Anatomy of Commodity Futures Risk Premia, *Journal of Finance* 69, 453-482.

Tang, Ke, and Wei Xiong, 2012, Index investment and financialization of commodities, *Financial Analyst Journal* 68, 54-74.

Working, Holbrook 1949, The theory of the price of storage, *American Economic Review* 39, 1254–1262.

Yang, Fan, 2013, Investment Shocks and the Commodity Basis Spread, *Journal of Financial Economics* 110, 164-184.

Figure 1: The Relationship between the Inventory and Convenience Yield

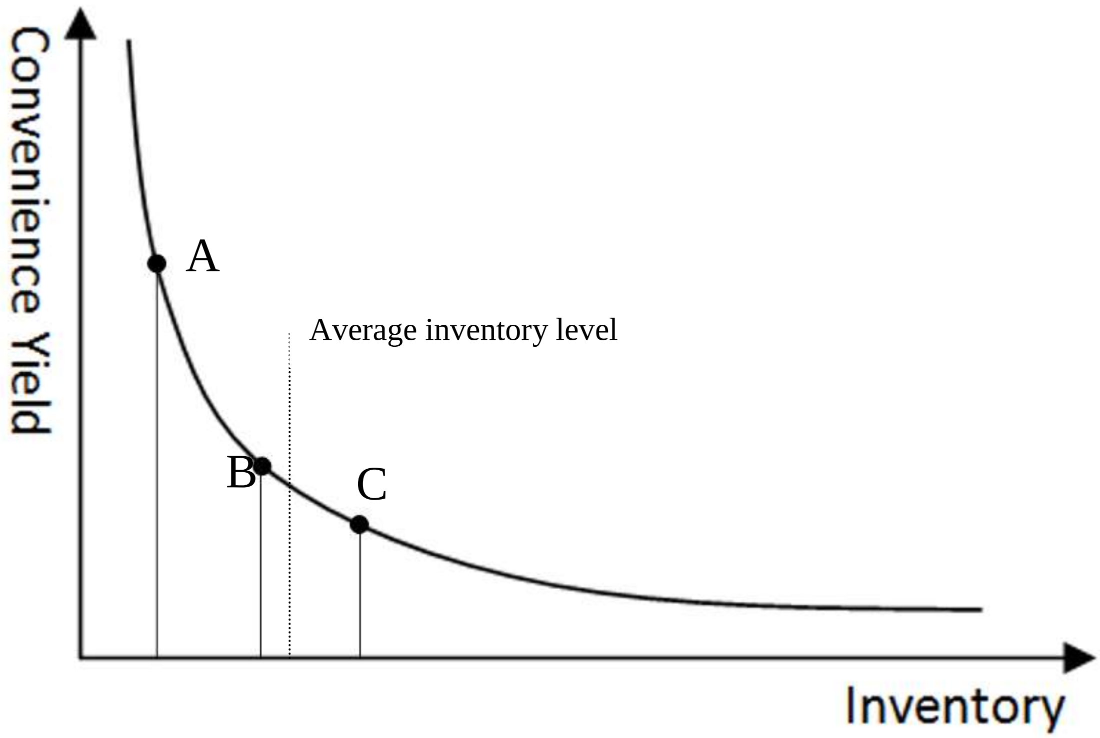

Table 1: Summary Statistics

This table reports the distribution of main variables. *TradtBasis* is the traditional basis, defined as the log-difference of the prices of the nearest and second-nearest futures contracts, scaled by their maturity time difference (in annualized number). *RelatBasis* is the relative basis, defined as the spread between a short-term and a long-term (or mid-term) *TradtBasis*. *ResidBasis* is the residual basis, defined as the sum of intercept and residual term of cross-sectional regression of *TradtBasis* on *RelatBasis*. *Momentum* is calculated as the cumulative past twelve month returns. *BasisMom* is the difference between momentum in a first- and second-nearby futures strategy. Smooth hedging pressure (*SHP*) is defined as the average netlong positions of commercials in the past one year. The sample is from January 1979 to June 2018 with a monthly frequency.

Panel A reports the time-series average of *TradtBasis*, *RelatBasis*, and return across 24 commodities. The panel reports the mean value and standard deviation of each variable, sorted into five commodity categories including energies, metals, softs, grains, and live stocks. Panel B reports the time-series average of cross-sectional correlations between main variables. Panel C reports the cross-sectional average autocorrelations (up to 6 lags) of 24 commodities of *TradtBasis*, *RelatBasis*, and *ResidBasis* variables (bold if significant at the 5% level).

Panel A: The time-series average of traditional basis, relative basis, and return across 24 commodities

	traditional basis		relative basis		return	
commodity name	mean	standard deviation	mean	standard deviation	mean	standard deviation
Crude Oil	0.39%	22.97%	-1.24%	10.48%	0.69%	9.47%
Heating Oil	0.46%	27.18%	0.77%	16.28%	0.61%	8.59%
Natural Gas	-20.51%	65.42%	-6.65%	52.06%	-0.55%	13.65%
Gold	-4.78%	3.95%	-0.01%	0.65%	0.10%	5.36%
Silver	-5.47%	5.50%	-0.45%	3.49%	0.18%	9.52%
Copper	0.98%	12.72%	-0.49%	5.12%	0.49%	7.69%
Platinum	-1.78%	5.35%	0.32%	5.22%	0.27%	7.46%
Palladium	-0.73%	5.62%	1.42%	16.00%	0.69%	9.05%
Cocoa	-5.89%	11.12%	0.22%	7.23%	-0.13%	8.50%
Coffee	-5.20%	16.53%	0.29%	8.30%	0.11%	10.62%
Orange Juice	-3.54%	17.00%	0.37%	11.41%	0.17%	8.86%
Sugar	-4.16%	26.12%	-4.30%	21.04%	0.02%	11.38%
Lumber	-12.98%	27.89%	-4.06%	21.30%	-0.51%	8.78%
Cotton	-3.00%	26.17%	-0.80%	27.66%	0.22%	7.40%
Soybean Oil	-5.28%	12.23%	-0.51%	6.67%	-0.15%	7.31%
Soybeans	-0.85%	21.06%	-2.01%	16.93%	0.18%	6.80%
Corn	-9.13%	18.86%	-1.67%	17.00%	-0.36%	7.26%
Wheat	-8.13%	19.19%	-4.14%	17.04%	-0.48%	7.57%
Oats	-6.76%	26.72%	-1.17%	18.43%	-0.03%	9.66%
Soybean Meal	5.39%	30.50%	2.01%	22.44%	0.74%	7.76%
Rough Rice	-11.73%	20.60%	-3.49%	23.80%	-0.56%	7.67%
Feeder Cattle	1.31%	14.88%	0.50%	12.60%	0.20%	4.30%
Live Cattle	0.55%	21.14%	-0.79%	22.95%	0.28%	4.38%
Lean Hogs	-8.91%	50.58%	-6.63%	55.93%	0.10%	7.49%
average	-4.57%	21.22%	-1.36%	17.50%	0.09%	8.19%

Panel B: Correlation

	RelatBasis	TradtBasis	ResidBasis	Momentum	BasisMom	SHP
RelatBasis	1.00	0.51	0.00	0.01	0.21	0.03
TradtBasis		1.00	0.72	0.35	0.36	0.15
ResidBasis			1.00	0.40	0.29	0.16
Momentum				1.00	0.27	0.31
BasisMom					1.00	0.03
SHP						1.00

Panel C: Autocorrelation

	RelatBasis	TradtBasis	ResidBasis
lag1	**0.33**	**0.68**	**0.71**
lag2	0.07	**0.51**	**0.56**
lag3	-0.02	**0.41**	**0.46**
lag4	-0.03	**0.33**	**0.38**
lag5	-0.03	**0.31**	**0.34**
lag6	0.00	**0.30**	**0.31**

Table 2: Return Predictability of Relative Basis: the Baseline Regression

This table examines the return predictability of the relative basis using Fama-Macbeth regression. The regression model is as follows:

$$ret_{i,t+1}(ret_{i,q+1}) = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}ResidBasis_{i,t}(\ or\ TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

Where $ret_{i,t+1}$ is the commodity cumulative excess returns in month $t+1$ and $ret_{i,q+1}$ is the commodity cumulative excess returns in quarter $q+1$. The main explanatory variables are *RelatBasis*, *ResidBasis,* and *TradtBasis*. *TradtBasis* is the traditional basis, defined as the log-difference of the prices of the nearest and second-nearest futures contracts, scaled by their maturity time difference (in annualized number). *RelatBasis* is the relative basis, defined as the spread between a short-term and a long-term (or mid-term) *TradtBasis*. *ResidBasis* is the residual basis, defined as the sum of intercept and residual term of cross-sectional regression of *TradtBasis* on *RelatBasis*. Control variables include momentum and BasisMom. *Momentum* is calculated as the cumulative past twelve month returns. *BasisMom* is the difference between momentum in a first- and second-nearby futures strategy. The sample is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Next one-month return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RelatBasis_{i,t}$	0.018			0.018			0.018	0.025
	(3.21)			*(2.61)*			*(2.46)*	*(2.91)*
$ResidBasis_{i,t}$		0.011			-0.009		-0.012	
		(1.60)			*(-1.18)*		*(-1.49)*	
$TradtBasis_{i,t}$			0.013			0.000		-0.012
			(2.53)			*(-0.01)*		*(-1.49)*
$Momentum_{i,t}$				0.015	0.014	0.013	0.018	0.018
				(3.20)	*(2.57)*	*(2.60)*	*(3.39)*	*(3.39)*
$BasisMom_{i,t}$				0.042	0.060	0.053	0.052	0.052
				(2.34)	*(3.08)*	*(2.99)*	*(2.69)*	*(2.69)*
Adj R^2	2.2%	3.3%	3.2%	9.9%	11.1%	10.7%	13.0%	13.0%

Panel B: Next one-quarter return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RelatBasis_{i,t}$	0.070			0.059			0.061	0.058
	(6.34)			*(4.75)*			*(4.73)*	*(3.48)*
$ResidBasis_{i,t}$		0.043			0.010		0.012	
		(2.30)			*(0.57)*		*(0.64)*	
$TradtBasis_{i,t}$			0.058			0.038		0.012
			(4.19)			*(2.61)*		*(0.64)*
$Momentum_{i,t}$				0.025	0.018	0.012	0.024	0.024
				(1.80)	*(1.27)*	*(0.88)*	*(1.65)*	*(1.65)*
$BasisMom_{i,t}$				0.111	0.137	0.103	0.105	0.105
				(2.41)	*(3.05)*	*(2.34)*	*(2.38)*	*(2.38)*
Adj R^2	2.5%	3.3%	3.4%	10.5%	11.1%	10.7%	13.5%	13.5%

Table 3: Return Predictability of Relative Basis for Financial Futures

This table examines the return predictability of the relative basis using 19 financial futures. The Fama-MacBeth regression model is as follows:

$$ret_{i,t+1}(ret_{i,q+1}) = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}ResidBasis_{i,t}(\,or\;TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

Where $ret_{i,t+1}$ is the commodity cumulative excess returns in month $t+1$ and $ret_{i,q+1}$ is the commodity cumulative excess returns in quarter $q+1$. The main explanatory variables are *RelatBasis*, *ResidBasis,* and *TradtBasis*. Control variables include momentum and BasisMom. The sample is from January 1993 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Next one-month return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RelatBasis_{i,t}$	-0.013			0.046			0.078	0.005
	(-0.11)			*(0.42)*			*(0.76)*	*(0.04)*
$ResidBasis_{i,t}$		0.115			0.071		0.073	
		(4.20)			*(2.35)*		*(2.51)*	
$TradtBasis_{i,t}$			0.092			0.068		0.073
			(3.04)			*(2.21)*		*(2.51)*
$Momentum_{i,t}$				0.034	0.030	0.028	0.031	0.031
				(4.54)	*(3.24)*	*(3.10)*	*(3.82)*	*(3.82)*
$BasisMom_{i,t}$				0.407	0.241	0.149	0.241	0.241
				(2.61)	*(1.87)*	*(1.01)*	*(1.37)*	*(1.37)*
Adj R^2	2.0%	3.6%	3.8%	21.7%	25.6%	26.9%	26.3%	26.3%

Panel B: Next one-quarter return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RelatBasis_{i,t}$	-0.115			-0.280			-0.188	-0.270
	(-0.45)			*(-0.97)*			*(-0.69)*	*(-0.88)*
$ResidBasis_{i,t}$		0.332			0.210		0.218	
		(4.53)			*(2.49)*		*(2.68)*	
$TradtBasis_{i,t}$			0.307			0.196		0.218
			(4.04)			*(2.58)*		*(2.68)*
$Momentum_{i,t}$				0.098	0.089	0.090	0.099	0.099
				(4.98)	*(4.01)*	*(4.02)*	*(4.70)*	*(4.70)*
$BasisMom_{i,t}$				1.011	0.863	0.712	0.779	0.779
				(2.31)	*(2.33)*	*(1.95)*	*(1.67)*	*(1.67)*
Adj R^2	2.0%	5.2%	5.3%	23.2%	27.7%	28.0%	30.3%	30.3%

Table 4: Return Predictability of Standardized Relative Basis

This table examines the return predictability based on standardized basis measures. The Fama-MacBeth regression model is as follows:

$$ret_{i,t+1}(ret_{i,q+1}) = a_t + b_{1,t}RelatBasis^{\mathrm{adj}}_{\mathrm{i,t}} + b_{2,t}ResidBasis^{\mathrm{adj}}_{\mathrm{i,t}}\left(or\ TradtBasis^{\mathrm{adj}}_{\mathrm{i,t}}\right)$$
$$+\lambda_t Controls^{\mathrm{adj}}_{\mathrm{i,t}} + \varepsilon_{i,t}$$

Where $ret_{i,t+1}$ is the cumulative futures returns in month $t+1$ and $ret_{i,q+1}$ is the cumulative futures returns in quarter $q+1$. The main explanatory variables are the standardized $RelatBasis^{adj}$, $ResidBasis^{adj}$, and $TradtBasis^{adj}$. Control variables include standardized $momentum^{adj}$ and standardized $BasisMom^{adj}$. For each month, we standardize the variables by subtracting their cross-sectional mean and then dividing their cross-sectional standard deviation. For each panel, the left three columns report the regression results of 24 commodities and the right three columns report the results of 19 financial futures. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Next one-month return as the dependent variable

	commodities futures			financial futures		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathrm{RelatBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0034	0.0034	0.0044	-0.0003	-0.0000	-0.0002
	(3.22)	*(3.09)*	*(2.74)*	*(-0.51)*	*(-0.04)*	*(-0.38)*
$\mathrm{ResidBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$		-0.0020			0.0013	
		(-1.84)			*(2.72)*	
$\mathrm{TradtBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$			-0.0025			0.0013
			(-1.58)			*(1.91)*
$\mathrm{Momentum}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0039	0.0049	0.0049	0.0032	0.0030	0.0030
	(3.13)	*(3.42)*	*(3.42)*	*(3.53)*	*(3.12)*	*(3.12)*
$\mathrm{BasisMom}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0024	0.0027	0.0027	0.0021	0.0011	0.0011
	(2.03)	*(2.36)*	*(2.36)*	*(3.14)*	*(1.45)*	*(1.45)*
Adj R^2	9.9%	13.0%	13.0%	21.7%	26.4%	26.4%

Panel B: Next one-quarter return as the dependent variable

	commodities futures			financial futures		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathrm{RelatBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0104	0.0108	0.0099	-0.0004	-0.0000	0.0001
	(5.14)	*(5.17)*	*(3.19)*	*(-0.27)*	*(-0.01)*	*(0.07)*
$\mathrm{ResidBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$		0.0018			0.0039	
		(0.69)			*(2.79)*	
$\mathrm{TradtBasis}^{\mathrm{adj}}_{\mathrm{i,t}}$			0.0019			0.0036
			(0.53)			*(2.17)*
$\mathrm{Momentum}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0062	0.0059	0.0059	0.0085	0.0088	0.0088
	(1.69)	*(1.50)*	*(1.50)*	*(3.48)*	*(3.50)*	*(3.50)*
$\mathrm{BasisMom}^{\mathrm{adj}}_{\mathrm{i,t}}$	0.0065	0.0061	0.0061	0.0047	0.0023	0.0023
	(2.17)	*(2.10)*	*(2.10)*	*(2.28)*	*(1.01)*	*(1.01)*
Adj R^2	10.5%	13.5%	13.5%	23.2%	30.3%	30.3%

Table 5: Portfolio Sorting Based on Relative Basis, Residual Basis, and Traditional Basis

This table examines the return predictability of different basis measures using portfolio sorting. The sorting variables include *RelatBasis*, *ResidBasis,* and *TradtBasis*. We form three equally-weighted portfolios every month based on sorting variables. At the end of month t, commodities are sorted into three groups based on their sorting variables over month t, from the lowest (P1) to the highest (P3). In month $t + 1$ to t+N, the average cumulative returns of each portfolio and return difference (P3-P1) are reported (in percent). The holding period is either next month (N=1) or next quarter (N=3). We report the cumulative return difference of P3 and P1, and associated t-statistics. Panel A reports the portfolio sorting results of 24 commodities from January 1979 to June 2018. Panel B reports the portfolio sorting results of 19 financial futures from January 1993 to June 2018. The t-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: The commodity futures markets

	P1	P2	P3	P3-P1	
	mean	mean	mean	mean	t-statistic
Sorted by the relative basis (RelatBasis)					
next month	-0.21%	-0.04%	0.56%	0.76%	*(3.83)*
next quarter	-0.93%	-0.03%	1.87%	2.79%	*(6.83)*
Sorted by the residual basis (ResidBasis)					
next month	-0.22%	0.07%	0.44%	0.66%	*(3.00)*
next quarter	-0.63%	0.01%	1.53%	2.16%	*(3.57)*
Sorted by the traditional basis(TradtBasis)					
next month	-0.25%	-0.01%	0.56%	0.81%	*(3.55)*
next quarter	-1.09%	0.02%	1.96%	3.05%	*(5.40)*

Panel B: The financial futures markets

	P1	P2	P3	P3-P1	
	mean	mean	mean	mean	t-statistic
Sorted by the relative basis (RelatBasis)					
next month	0.16%	0.11%	0.25%	0.09%	*(0.64)*
next quarter	0.49%	0.40%	0.69%	0.20%	*(0.69)*
Sorted by the residual basis (ResidBasis)					
next month	-0.09%	0.17%	0.41%	0.50%	*(4.37)*
next quarter	-0.25%	0.60%	1.14%	1.39%	*(4.30)*
Sorted by the traditional basis(TradtBasis)					
next month	-0.05%	0.21%	0.35%	0.41%	*(2.76)*
next quarter	-0.23%	0.59%	1.14%	1.37%	*(3.46)*

Table 6: Strategy Portfolios Returns Based on Relative Basis, Residual Basis, and Traditional Basis

This table runs the regression of one strategy portfolio return on other strategy portfolio returns and examines which strategy portfolio returns are the most prominent factors in the commodity pricing. The sample is sorted into three groups based on sorting variables, including *RelatBasis*, *ResidBasis*, *TradtBasis*, *momentum*, and *BasisMom*. The strategy portfolio return is calculated as the return spread between the low and high group (P3-P1) of sorting variables. The market factor *MKTRet* is calculated as the equally weighted return of 24 commodities. The sample contains 24 commodities from January 1979 to June 2018.

	$RelatBasisRet_t$	$RelatBasisRet_t$	$RelatBasisRet_t$	$ResidBasisRet_t$	$ResidBasisRet_t$	$TradtBasisRet_t$	$TradtBasisRet_t$
Intercept	0.664	0.663	0.585	0.106	0.106	0.261	0.061
	(3.29)	*(3.29)*	*(3.04)*	*(0.56)*	*(0.55)*	*(1.30)*	*(0.31)*
$RelatBasisRet_t$					0.000		0.301
					(-0.00)		*(6.86)*
$ResidBasisRet_t$		0.000					
		(-0.00)					
$TradtBasisRet_t$			0.303				
			(6.86)				
$MKTRet_t$	0.132	0.131	0.146	-0.084	-0.084	-0.046	-0.086
	(2.32)	*(2.31)*	*(2.69)*	*(-1.56)*	*(-1.55)*	*(-0.81)*	*(-1.58)*
$MomRet_t$	-0.105	-0.105	-0.196	0.371	0.371	0.302	0.334
	(-2.69)	*(-2.43)*	*(-4.97)*	*(10.06)*	*(9.97)*	*(7.77)*	*(8.92)*
$BasisMomRet_t$	0.185	0.185	0.099	0.229	0.229	0.283	0.227
	(4.35)	*(4.21)*	*(2.34)*	*(5.73)*	*(5.61)*	*(6.69)*	*(5.52)*
Adj R^2	5.1%	4.9%	13.6%	24.3%	24.2%	20.4%	27.5%

Table 7: Strategy Portfolio Returns among Different Business Cycles

This table examines the return predictability of relative basis among different business cycles. The business cycle index is obtained from Chicago Fed. The full sample is equally split into two subsamples: contraction and expansion periods. Panel A conducts the portfolio sorting based on the relative basis and reports the difference in return spreads of low and high relative basis groups between two periods. Panel B runs the regression of the relative basis strategy portfolio return on other strategy portfolio returns among two business cycle periods. The sample is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Portfolio sorting based on relative basis

		P1	P3	P3-P1
contraction	next month	-0.13%	0.31%	0.44%
				(*1.56*)
	next quarter	-1.03%	1.05%	2.07%
				(*4.06*)
expansion	next month	-0.28%	0.80%	1.09%
				(*3.81*)
	next quarter	-0.83%	2.68%	3.51%
				(*7.05*)
Expansion - contraction	next month			0.65%
				(*1.63*)
	next quarter			1.44%
				(*2.02*)

Panel B: Strategy portfolio returns based on relative basis

monthly return of the relative basis strategy portfolio			quarterly return of the relative basis strategy portfolio		
	contraction	expansion		contraction	expansion
Intercept	0.329	0.972	Intercept	1.547	3.074
	(1.17)	*(3.34)*		*(2.84)*	*(5.73)*
$MKTRet_t$	0.106	0.193	$MKTRet_q$	0.195	0.095
	(1.52)	*(1.96)*		*(2.78)*	*(1.00)*
$MomRet_t$	-0.131	-0.075	$MomRet_q$	0.013	-0.049
	(-2.45)	*(-1.31)*		*(0.23)*	*(-0.86)*
$BasisMomRet_t$	0.189	0.189	$BasisMomRet_q$	0.147	0.202
	(3.38)	*(2.93)*		*(2.18)*	*(3.24)*
Adj R^2	5.6%	4.1%	Adj R^2	4.9%	3.3%

Table 8: Asymmetric Pattern of Return Predictability of Relative Basis

This table examines the asymmetric pattern of return predictability of relative basis. The Fama-MacBeth regression model is as follows:

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis\_pos_{i,t} + b_{2,t}RelatBasis\_neg_{i,t} \\ +b_{3,t}ResidBasis_{i,t}(\ or\ TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

Where ret_{t+1} is the commodity cumulative excess returns in month $t+1$. The main explanatory variables are *RelatBasis_pos*, *RelatBasis_neg*, *ResidBasis*, and *TradtBasis*. We define *RelatBasis_pos* = *RelatBasis* if *RelatBasis* >=0, otherwise *RelatBasis_pos*=0; *RelatBasis_neg* = *RelatBasis* if *RelatBasis* <0, otherwise *RelatBasis_neg*=0. Control variables include momentum and BasisMom. The sample is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

	(1)	(2)	(3)	(4)
$RelatBasis\_pos_{i,t}$	0.043	0.044	0.045	0.056
	(2.83)	*(2.44)*	*(2.56)*	*(3.27)*
$RelatBasis\_neg_{i,t}$	0.018	0.009	0.007	0.018
	(1.41)	*(0.73)*	*(0.56)*	*(1.22)*
$ResidBasis_{i,t}$			-0.010	
			(-1.21)	
$TradtBasis_{i,t}$				-0.010
				(-1.21)
$Momentum_{i,t}$		0.014	0.017	0.017
		(3.01)	*(3.30)*	*(3.30)*
$BasisMom_{i,t}$		0.046	0.053	0.053
		(2.45)	*(2.70)*	*(2.70)*
Adj R^2	4.1%	11.7%	14.6%	14.6%

Table 9: Samples Partitioned on the Absolute Value of RelatBasis

This table examines the return predictability of relative basis among two subsamples partitioned on the absolute value of the relative basis *abs(RelatBasis)*. The full sample is equally split into two subsamples based on *abs(RelatBasis)* at the end of month t: high and low abs(RelatBasis) samples. The table conducts the baseline regression within each subsample. The dependent variables are next month and next quarter commodity returns. The explanatory variables are *RelatBasis* and *ResidBasis (or TradtBasis)*. Control variables include momentum and BasisMom. For each month, we standardize the variables by subtracting their cross-sectional mean and then dividing their cross-sectional standard deviation. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

	high abs(RelatBasis)			low abs(RelatBasis)		
next month	(1)	(2)	(3)	(4)	(5)	(6)
$\text{RelatBasis}^{adj}_{i,t}$	0.0052	0.0048	0.0091	0.0023	0.0011	0.0012
	(3.37)	*(2.91)*	*(3.69)*	*(1.73)*	*(0.73)*	*(0.72)*
$\text{ResidBasis}^{adj}_{i,t}$		-0.0064			0.0013	
		(-3.56)			*(0.65)*	
$\text{TradtBasis}^{adj}_{i,t}$			-0.0086			0.0011
			(-3.47)			*(0.53)*
controls	yes	yes	yes	yes	yes	yes
Adj R^2	12.1%	15.7%	15.7%	15.2%	20.6%	20.6%
next quarter	(1)	(2)	(3)	(4)	(5)	(6)
$\text{RelatBasis}^{adj}_{i,t}$	0.0139	0.0128	0.0130	0.0053	0.0031	0.0030
	(4.78)	*(4.00)*	*(3.18)*	*(1.86)*	*(0.98)*	*(0.96)*
$\text{ResidBasis}^{adj}_{i,t}$		-0.0017			0.0047	
		(-0.56)			*(1.38)*	
$\text{TradtBasis}^{adj}_{i,t}$			-0.0021			0.0044
			(-0.46)			*(1.27)*
controls	yes	yes	yes	yes	yes	yes
Adj R^2	13.4%	14.7%	14.7%	15.9%	22.9%	22.9%

Table 10: Spread Return Predictability of Relative Basis

This table examines the spread return predictability of the relative basis using Fama-MacBeth regression. The regression model is as follows:

$$ret_{spread_{i,t+1}} = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}ResidBasis_{i,t}(\,or\;TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

Where $ret_{spread_{t+1}}$ is the difference in returns between a first- and second-nearby futures in month t + 1. The main explanatory variables are *RelatBasis*, *ResidBasis,* and *TradtBasis*. *TradtBasis* is the traditional basis, defined as the log-difference of the prices of the nearest and second-nearest futures contracts, scaled by their maturity time difference (in annualized number). *RelatBasis* is the relative basis, defined as the spread between a short-term and a long-term (or mid-term) *TradtBasis*. *ResidBasis* is the residual basis, defined as the sum of intercept and residual term of cross-sectional regression of *TradtBasis* on *RelatBasis*. Control variables include momentum and BasisMom. *Momentum* is calculated as the cumulative past twelve month returns. *BasisMom* is the difference between momentum in a first- and second-nearby futures strategy. The sample is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RelatBasis_{i,t}$	0.004			0.003			0.004	0.004
	(2.14)			*(1.94)*			*(1.94)*	*(1.84)*
$ResidBasis_{i,t}$		-0.001			0.000		-0.001	
		(-0.35)			*(-0.07)*		*(-0.59)*	
$TradtBasis_{i,t}$			0.001			0.001		-0.001
			(0.43)			*(0.71)*		*(-0.59)*
$Momentum_{i,t}$				-0.002	-0.003	-0.003	-0.002	-0.002
				(-2.44)	*(-2.59)*	*(-2.84)*	*(-1.88)*	*(-1.88)*
$BasisMom_{i,t}$				0.014	0.017	0.015	0.013	0.013
				(3.01)	*(3.53)*	*3.22*	*(2.79)*	*(2.79)*
Adj R^2	8.8%	5.7%	7.8%	16.2%	13.3%	15.0%	21.7%	21.7%

Table 11: Inventory and Relative Basis

Panel A runs the cross-sectional regressions of commodity relative basis (or residual basis or traditional basis) on inventory and the inverse of inventory. The Fama-MacBeth regression model is as follows:

$$RelatBasis_{i,t}(or\ TradtBasis_{i,t}\ or\ ResidBasis_{i,t}) \\ = a_t + b_{1,t}Inventory_{i,t} + b_{2,t}(1/\ Inventory_{i,t}) + \varepsilon_{i,t}$$

Where *TradtBasis* is the traditional basis, defined as the log-difference of the prices of the nearest and second-nearest futures contracts, scaled by their maturity time difference (in annualized number). *RelatBasis* is the relative basis, defined as the spread between a short-term and a long-term (or mid-term) *TradtBasis*. *ResidBasis* is the residual basis, defined as the sum of intercept and residual term of cross-sectional regression of *TradtBasis* on *RelatBasis*. *Inventory* is the ratio of inventory level standardized by its past 12 moving average.

Panel B decomposes the relative basis into two components using the regression equation in panel A and examines the return predictability of each component using the following regression:

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis_{INV\,i,t} + b_{2,t}RelatBasis_{NOINV\,i,t} \\ + b_{3,t}ResidBasis_{i,t}(\ or\ TradtBasis_{i,t}) + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

Where $RelatBasis_{INV}$ is the fitted value of the regression equation and $RelatBasis_{NOINVi,t}$ is the residual term. Control variables include momentum and BasisMom. The sample is from January 1993 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Relative basis, inventory, and scarcity of inventory

	$RelatBasis_{i,t}$	$ResidBasis_{i,t}$	$TradtBasis_{i,t}$
$Inventory_{i,t}$	0.030	-0.071	-0.073
	(1.28)	*(-3.70)*	*(-2.60)*
$1/Inventory_{i,t}$	0.097	0.001	0.058
	(3.98)	*(0.08)*	*(2.08)*
Adj R^2	0.2%	0.2%	0.9%

Panel B: The decomposition of relative basis

	(1)	(2)	(3)	(4)
$RelatBasis_{INVi,t}$	0.159	0.151	0.143	0.153
	(2.79)	*(2.45)*	*(2.37)*	*(2.53)*
$RelatBasis_{NOINVi,t}$	0.008	0.012	0.009	0.018
	(1.13)	*(1.45)*	*(0.94)*	*(1.65)*
$ResidBasis_{i,t}$			-0.014	
			(-1.18)	
$TradtBasis_{i,t}$				-0.014
				(-1.18)
$Momentum_{i,t}$		0.013	0.016	0.016
		(2.78)	*(2.91)*	*(2.91)*
$BasisMom_{i,t}$		0.018	0.033	0.033
		(0.70)	*(1.02)*	*(1.02)*
Adj R^2	4.6%	11.7%	15.6%	15.6%

Appendix A1: Variables Definition

$ret_{i,t}$	The excess return $ret_{i,t}$ on commodity *i* in month *t*. $ret_{i,t} = \frac{F_i(t,T) - F_i(t-1,T)}{F_i(t-1,T)}$ where $F_i(t-1,T)$ is the futures price at the end of month *t*-1 on the nearest contract with expiration date *T* and $F_i(t,T)$ is the price of the same contract at the end of month t.
$TradtBasis_{i,t}(T1,T2)$	$\text{TradtBasis}_{i,t} = \frac{\log(F_i(t,T1)) - \log(F_i(t,T2))}{T2-T1}$ Where $F_i(t,T1)$ and $F_i(t,T2)$ are short-term and long-term futures prices at the end of month *t* with expiration date *T1* and *T2*.
$RelatBasis_{i,t}$	The spread between a short-term and a long-term (or mid-term) traditional basis or slope of traditional basis. $RelatBasis_{i,t} = Tradt\text{Basis}_{i,t}(T1,T2) - Tradt\text{Basis}_{i,t}(T2,T3)$
$ResidBasis_{i,t}$	*ResidBasis* is the residual basis, defined as the sum of intercept and residual term of cross-sectional regression of *TradtBasis* on *RelatBasis*. The cross-sectional regression is as follows: $\text{TradtBasis}_{i,t} = a_t + b_t * \text{RelatBasis}_{i,t} + \varepsilon_{i,t}$, $RelatBasis_{i,t} = \alpha_t + \varepsilon_{i,t}$.
$Inventory_{i,t}$	$\text{Inventory}_{i,t} = RawInventory_i(t)/RawInventory_i(t-1{:}\,t-12)$ where $RawInventory_{i,t}$ is the inventory level of commodity *i* in month *t*; $RawInventory_i(t-1{:}\,t-12)$ is the inventory moving average of commodity *i* in the previous 12 months.
$Scarcity_{i,t}$	Scarcity of inventory=1/$\text{RawInventory}_{i,t}$
$SHP_{i,t}$	The smoothed hedging pressure (*SHP*) is defined as the average of the net short position (short position minus long position) of the commercial traders in the Commitments of Traders (COT) report from CFTC in the past one year, scaled by the commodity's most recent open interest.
$Momentum_{i,t}$	The cumulative past twelve month returns of firm *i* from *t*-11 to *t*.
$BasisMom_{i,t}$	The difference between momentum in a first- and second-nearby futures strategy in Boons and Prado (2019).

Appendix A2: Volume and Open Interest of Individual Commodity Futures Contract

This table reports the volume and open interest of nearest, second-nearest, and third-nearest contracts across 24 commodities. The 24 commodities are sorted into five commodity categories including energies, metals, softs, grains, and live stocks. The table reports the time-series average of volume (Vol1, Vol2, and Vol3) and open interest (OI1, OI2, and OI3) of each commodity. The sample is from January 1979 to June 2018 with a monthly frequency.

commodity name	Vol1	Vol2	Vol3	Vol3/Vol2	OI1	OI2	OI3	OI3/OI2
Crude Oil	97750	38485	15839	41%	194315	104544	64505	62%
Heating Oil	22079	6939	3469	50%	54531	27762	18439	66%
Natural Gas	42686	13795	8247	60%	118501	74686	52883	71%
Gold	21236	26243	5166	20%	63285	85078	30777	36%
Silver	10702	8633	634	7%	36004	34943	9887	28%
Copper	7530	6368	593	9%	33151	33783	9551	28%
Platinum	2167	1198	64	5%	13821	8970	1004	11%
Palladium	750	461	23	5%	7544	4560	366	8%
Cocoa	3964	5175	1420	27%	21502	33876	16045	47%
Coffee	4592	5039	1249	25%	22924	30048	10781	36%
Orange Juice	1141	1058	181	17%	5999	8127	2500	31%
Sugar	17586	17471	6120	35%	102121	108418	51777	48%
Lumber	701	640	129	20%	2319	2329	729	31%
Cotton	3747	5432	2400	44%	20950	37648	19550	52%
Soybean Oil	6213	7082	2398	34%	28970	53951	28122	52%
Soybeans	14983	14444	4697	33%	58592	85014	41601	49%
Corn	20799	17625	6797	39%	157021	202054	100788	50%
Wheat	6971	7476	2182	29%	56440	68995	29269	42%
Oats	622	595	128	22%	3584	4246	1469	35%
Soybean Meal	6340	6368	2310	36%	24925	42739	24381	57%
Rough Rice	231	247	40	16%	2691	3751	1062	28%
Feeder Cattle	1447	1348	703	52%	6851	6639	3996	60%
Live Cattle	10457	9573	3953	41%	45596	53615	27477	51%
Lean Hogs	5943	6713	2756	41%	22937	31900	16481	52%
average				30%				43%

Appendix A3: Panel Regression: Return Predictability of Relative Basis

This table examines the return predictability based on different basis measures using the panel regression with fixed effects. The dependent variables, explanatory variables, and control variables are the same as table 2. In addition, the panel regression controls the time and commodity fixed effects. The sample period is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts and are displayed in parentheses below. The associated standard errors are clustered by time and commodity.

Panel A: Next one-month return as the dependent variable

	1	2	3	4	5	6	7	8
$RelatBasis_{i,t}$	0.013			0.011			0.011	0.021
	(3.23)			*(2.60)*			*(2.53)*	*(3.68)*
$ResidBasis_{i,t}$		0.001			-0.008		-0.007	
		(0.13)			*(-1.24)*		*(-1.09)*	
$TradtBasis_{i,t}$			0.003			-0.004		-0.016
			(0.61)			*(-1.11)*		*(-3.07)*
$Momentum_{i,t}$				0.005	0.007	0.006	0.007	0.009
				(2.21)	*(2.35)*	*(2.25)*	*(2.40)*	*(3.09)*
$BasisMom_{i,t}$				0.039	0.047	0.048	0.041	0.045
				(4.13)	*(5.30)*	*(5.35)*	*(4.61)*	*(5.02)*
Time and commodity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	13.4%	13.3%	13.3%	13.5%	13.5%	13.5%	13.5%	13.6%

Panel B: Next one-quarter return as the dependent variable

	1	2	3	4	5	6	7	8
$RelatBasis_{i,t}$	0.056			0.048			0.049	0.048
	(3.95)			*(3.69)*			*(3.83)*	*(2.70)*
$ResidBasis_{i,t}$		0.023			0.009		0.013	
		(1.59)			*(0.61)*		*(0.86)*	
$TradtBasis_{i,t}$			0.037			0.028		0.000
			(3.86)			*(2.88)*		*(0.01)*
$Momentum_{i,t}$				0.005	0.001	-0.003	0.002	0.005
				(0.54)	*(0.12)*	*(-0.35)*	*(0.21)*	*(0.45)*
$BasisMom_{i,t}$				0.097	0.121	0.103	0.092	0.097
				(3.75)	*(5.28)*	*(4.45)*	*(3.90)*	*(4.05)*
Time and commodity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	14.9%	14.4%	14.6%	15.2%	14.7%	14.9%	15.2%	15.2%

Appendix A4: The Distribution of Traditional Basis, Relative Basis, and Return for Financial Futures

This table reports the time-series average of traditional basis, relative basis, and return across 19 financial futures. The table reports the mean value and standard deviation of each variable, sorted by the standard deviation of relative basis. The sample is from January 1993 to June 2018.

Financial Futures	traditional basis		relative basis		return	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
Dow Jones Industrial Mini-Sized	0.73%	1.84%	0.02%	0.39%	0.65%	3.72%
NASDAQ 100 Index, E-mini	-1.07%	2.01%	0.12%	0.31%	0.45%	6.63%
NIKKEI 225 Index	-0.45%	1.35%	-0.20%	1.37%	0.22%	5.90%
S&P 500 Index, E-mini	1.01%	1.23%	0.07%	0.18%	0.62%	4.30%
S&P 500 Index	-1.03%	2.41%	0.08%	0.24%	0.60%	4.09%
Federal Funds / 30-day	0.27%	1.26%	-0.01%	0.85%	0.03%	0.13%
Treasury Note, U.S., 2-year	0.89%	0.71%	0.26%	0.99%	0.08%	0.32%
Treasury Note, U.S., 5-year	1.81%	1.55%	0.86%	1.75%	0.19%	1.01%
Treasury Note, U.S., 10-year	2.83%	1.84%	0.49%	1.47%	0.34%	1.70%
Treasury Bonds, U.S., 30-year	2.36%	5.43%	0.22%	6.40%	0.38%	2.78%
Australian Dollar / U.S. Dollar	1.83%	1.57%	0.03%	0.21%	0.24%	3.39%
British Pound / U.S. Dollar	0.70%	1.14%	0.01%	0.18%	0.05%	2.43%
Canadian Dollar / U.S. Dollar	0.11%	0.99%	-0.01%	0.18%	0.03%	2.37%
Eurodollar, 3-month	0.41%	0.90%	-0.12%	0.60%	0.03%	0.19%
Euro FX	-0.51%	1.28%	0.05%	0.19%	0.02%	2.89%
Japanese Yen / U.S. Dollar	-2.61%	2.03%	0.07%	0.26%	-0.13%	3.15%
New Zealand Dollar / U.S. Dollar	2.25%	1.47%	0.00%	0.61%	0.26%	3.78%
U.S. Dollar Index	-0.59%	1.51%	-0.05%	0.31%	-0.02%	2.31%
Swiss Franc / U.S. Dollar	-1.78%	1.49%	0.04%	0.25%	0.03%	3.06%
average	0.38%	1.68%	0.10%	0.88%	0.21%	2.85%

Appendix A5: Return Predictability of Relative Basis Controlling Smooth Hedging Pressure

This table examines the return predictability based on different basis measures. The Fama-Macbeth regression model is similar to Table 2 using the next one-month return as the dependent variable, and the regression model controls additional variable smooth hedging pressure (*SHP*). *SHP* is calculated as the average of the net short position of the commercial traders in the past one year, scaled by the commodity's most recent open interest. Panel A and Panel B report the regression results of 24 commodities and 19 financial futures. The sample is from January 1993 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: The commodity futures markets

	(1)	(2)	(3)
$RelatBasis_{i,t}$	0.017	0.017	0.022
	(2.37)	*(2.27)*	*(2.23)*
$ResidBasis_{i,t}$		-0.011	
		(-1.15)	
$TradtBasis_{i,t}$			-0.011
			(-1.15)
$Momentum_{i,t}$	0.012	0.015	0.015
	(2.58)	*(2.84)*	*(2.84)*
$BasisMom_{i,t}$	0.054	0.070	0.070
	(2.59)	*(2.88)*	*(2.88)*
$SHP_{i,t}$	0.013	0.012	0.012
	(2.38)	*(2.31)*	*(2.31)*
Adj R^2	10.5%	13.9%	13.9%

Panel B: The financial futures markets

	(1)	(2)	(3)
$RelatBasis_{i,t}$	-0.011	0.033	-0.101
	(-0.11)	*(0.30)*	*(-0.94)*
$ResidBasis_{i,t}$		0.099	
		(3.67)	
$TradtBasis_{i,t}$			0.099
			(3.67)
$Momentum_{i,t}$	0.036	0.034	0.034
	(3.89)	*(3.39)*	*(3.39)*
$BasisMom_{i,t}$	0.529	0.391	0.391
	(3.02)	*(1.88)*	*(1.88)*
$SHP_{i,t}$	0.001	0.001	0.001
	(0.44)	*(0.22)*	*(0.22)*
Adj R^2	25.3%	27.3%	27.3%

Appendix A6: Return Predictability of Relative Basis for Commodities among Sub-sample Periods

This table examines the return predictability of relative basis using the subsample period of 1979-1999 and 2000-2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Next one-month return as the dependent variable

	1979-1999			2000-2018		
$RelatBasis_{i,t}$	0.014	0.017	0.029	0.023	0.018	0.020
	(1.71)	*(1.65)*	*(2.46)*	*(3.03)*	*(1.87)*	*(1.65)*
$ResidBasis_{i,t}$		-0.023			0.002	
		(-2.22)			*(0.18)*	
$TradtBasis_{i,t}$			-0.023			0.002
			(-2.22)			*(0.18)*
$Momentum_{i,t}$		0.025	0.025		0.010	0.010
		(3.15)	*(3.15)*		*(1.64)*	*(1.64)*
$BasisMom_{i,t}$		0.038	0.038		0.067	0.067
		(1.62)	*(1.62)*		*(2.24)*	*(2.24)*
Adj R^2	1.4%	12.7%	12.7%	3.0%	13.3%	13.3%

Panel B: Next one-quarter return as the dependent variable

	1979-1999			2000-2018		
$RelatBasis_{i,t}$	0.060	0.063	0.060	0.080	0.058	0.056
	(4.18)	*(3.53)*	*(2.38)*	*(4.92)*	*(3.13)*	*(2.63)*
$ResidBasis_{i,t}$		0.001			0.025	
		(0.02)			*(1.06)*	
$TradtBasis_{i,t}$			0.001			0.025
			(0.02)			*(1.06)*
$Momentum_{i,t}$		0.025	0.025		0.023	0.023
		(1.09)	*(1.09)*		*(1.42)*	*(1.42)*
$BasisMom_{i,t}$		0.091	0.091		0.122	0.122
		(1.60)	*(1.60)*		*(1.82)*	*(1.82)*
Adj R^2	1.8%	14.1%	14.1%	3.3%	12.9%	12.9%

Appendix A7: Alternative Construction Method of Relative Basis

This table examines the return predictability based on basis measures using an alternative construction method, where we adjust the maturities of the first, second, and third nearby contract of crude oil, heating oil, and natural gas, so that relative basis employs the same horizon for all commodities. The Fama-Macbeth regression model is similar to Table 2. The sample period is from January 1979 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: Next one-month return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{RelatBasis}_{i,t}$	0.018			0.018			0.018	0.024
	(2.92)			*(2.76)*			*(2.55)*	*(2.96)*
$\text{ResidBasis}_{i,t}$		0.010			-0.011		-0.014	
		(1.37)			*(-1.38)*		*(-1.71)*	
$\text{TradtBasis}_{i,t}$			0.010			-0.005		-0.014
			(1.96)			*(-0.86)*		*(-1.71)*
$\text{Momentum}_{i,t}$				0.015	0.015	0.015	0.019	0.019
				(3.14)	*(2.91)*	*(2.99)*	*(3.74)*	*(3.74)*
$\text{BasisMom}_{i,t}$				0.045	0.053	0.060	0.048	0.048
				(2.47)	*(2.95)*	*(3.34)*	*(2.64)*	*(2.64)*
Adj R^2	2.4%	3.3%	3.4%	9.9%	10.9%	11.0%	12.9%	12.9%

Panel B: Next one-quarter return as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{RelatBasis}_{i,t}$	0.068			0.062			0.061	0.066
	(5.41)			*(4.74)*			*(4.59)*	*(3.75)*
$\text{ResidBasis}_{i,t}$		0.035			-0.002		-0.001	
		(1.75)			*(-0.09)*		*(-0.05)*	
$\text{TradtBasis}_{i,t}$			0.052			0.028		-0.001
			(3.76)			*(1.98)*		*(-0.05)*
$\text{Momentum}_{i,t}$				0.025	0.024	0.017	0.030	0.030
				(1.83)	*(1.71)*	*(1.21)*	*(2.10)*	*(2.10)*
$\text{BasisMom}_{i,t}$				0.105	0.134	0.116	0.099	0.099
				(2.28)	*(3.05)*	*(2.67)*	*(2.27)*	*(2.27)*
Adj R^2	2.8%	3.5%	3.4%	10.7%	11.3%	10.9%	13.9%	13.9%

Appendix A8: Risk Factor Correlation

The sample is sorted into three groups based on sorting variables, including *RelatBasis*, *ResidBasis*, *TradtBasis*, *momentum*, and *BasisMom*. The strategy portfolio return is calculated as the return spread between the low and high group (P3-P1) of sorting variables. The market factor *MKTRet* is calculated as the equally weighted return of 24 commodities. The table reports the correlation between risk factors. The sample is from January 1979 to June 2018.

	RelatBasisRet	ResidBasisRet	TradtBasisRet	MKT Ret	MomRet	BasisMomRet
RelatBasisRet	1.00	0.00	0.28	0.10	-0.08	0.18
ResidBasisRet		1.00	0.78	0.00	0.44	0.29
TradtBasisRet			1.00	0.02	0.36	0.33
MKTRet				1.00	0.13	0.06
MomRet					1.00	0.16
BasisMomRet						1.00

Appendix A9: The Mean Reverting Attributes of Inventory

Panels of A and B investigate the mean-reverting pattern of inventory. The regression models are as follows:

$$\text{Model 1: Inventory}_{i,t} - \text{Inventory}_{i,t-1} = a_t + b_t\text{Inventory}_{i,t-1} + \varepsilon_{i,t}$$

$$\text{Model 2: } \text{Ln}(\text{RawInventory}_{i,t}) - \text{Ln}(\text{RawInventory}_{i,t-1}) = a_t + b_t\text{Ln}(\text{RawInventory}_{i,t-1}) + \varepsilon_{i,t}$$

Specifically, Panel A runs the time-series regression for each commodity, and Panel B runs the pool regression using the full sample controlling the commodity fixed effect.

Panel C investigates the relationship between change in inventory and the relative basis. The Fama-MacBeth regression model is as follows:

$$\begin{aligned} Inventory_{i,t+N} - Inventory_{i,t} = a_t + b_{1,t}RelatBasis_{i,t} \\ +b_{2,t}ResidBasis_{i,t}(\ or\ TradtBasis_{i,t}) + \varepsilon_{i,t} \end{aligned}$$

where $Inventory_{i,t+N}$ is the commodity inventory in month t+N (N=1, 2, 3) and $Inventory_{i,t}$ is the commodity inventory in month t. The explanatory variables are *RelatBasis*, *ResidBasis*, and *TradtBasis*. The sample is from January 1993 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

Panel A: time-series regression of each commodity

	Model 1			Model 2		
commodity name	inventory_{t-1}	*t*-stat	Adj R^2	$\text{Ln(RawInventory}_{t-1})$	*t*-stat	Adj R^2
Crude Oil	-0.135	*-4.64*	6.6%	-0.017	*-1.50*	0.7%
Heating Oil	-0.151	*-4.96*	7.5%	-0.083	*-3.60*	4.1%
Natural Gas	-0.139	*-4.76*	7.0%	-0.121	*-4.42*	6.1%
Gold	-0.131	*-4.62*	6.6%	-0.008	*-1.17*	0.5%
Silver	-0.066	*-3.16*	3.2%	-0.009	*-1.16*	0.4%
Copper	-0.083	*-3.61*	4.1%	-0.027	*-2.03*	1.3%
Platinum	-0.237	*-5.13*	12.0%	-0.025	*-1.37*	1.0%
Palladium	-0.119	*-3.49*	6.0%	-0.019	*-1.37*	1.0%
Cocoa	-0.105	*-4.10*	5.3%	-0.042	*-2.49*	2.0%
Coffee	-0.112	*-3.93*	5.9%	-0.067	*-6.69*	15.4%
Orange Juice	-0.117	*-4.25*	5.6%	-0.029	*-2.06*	1.4%
Sugar	-0.341	*-7.93*	17.2%	-0.335	*-7.81*	16.7%
Lumber	-0.431	*-6.72*	21.7%	-0.063	*-2.25*	3.0%
Cotton	-0.197	*-5.65*	9.8%	-0.138	*-4.65*	6.9%
Soybean Oil	-0.076	*-3.05*	3.7%	-0.021	*-1.65*	1.1%
Soybeans	-0.282	*-6.42*	13.9%	-0.235	*-5.66*	11.1%
Corn	-0.392	*-8.59*	19.6%	-0.368	*-8.26*	18.4%
Wheat	-0.141	*-4.05*	7.1%	-0.081	*-2.97*	4.0%
Oats	-0.327	*-7.67*	16.2%	-0.144	*-4.82*	7.1%
Soybean Meal	-0.698	*-10.78*	35.4%	-0.515	*-8.54*	25.6%
Rough Rice	-0.347	*-7.72*	17.3%	-0.336	*-7.56*	16.7%

Feeder Cattle	-0.169	*-5.42*	8.8%	-0.032	*-2.20*	1.6%
Live Cattle	-0.135	*-4.69*	6.8%	-0.056	*-3.14*	3.2%
Lean Hogs	-0.197	*-5.75*	9.8%	-0.067	*-3.38*	3.6%
average	-0.214	*-7.09*	10.7%	-0.118	*-4.21*	6.4%

Panel B: Pool regression with commodity fixed effect

	change in inventory	
Model 1	coefficient	*t*-statistic
Intercept	0.242	*13.96*
inventory_{t-1}	-0.238	*-29.80*
commodity fixed effect	*yes*	
Adj R^2	11.9%	
	change in log of raw inventory	
Model 2	coefficient	*t*-statistic
Intercept	0.577	*16.72*
$\text{Ln(RawInventory}_{t-1})$	-0.112	*-20.02*
commodity fixed effect	*yes*	
Adj R^2	5.8%	

Panel C: Change in inventory and relative basis

	N=1	N=2	N=3
$\text{RelatBasis}_{i,t}$	0.011	0.086	0.242
	(0.63)	*(3.61)*	*(6.32)*
$\text{ResidBasis}_{i,t}$	-0.067	-0.104	-0.075
	(-3.09)	*(-2.80)*	*(-1.59)*
Adj R^2	1.1%	0.7%	0.3%
$\text{RelatBasis}_{i,t}$	0.078	0.180	0.336
	(2.78)	(4.13)	(5.45)
$\text{TradtBasis}_{i,t}$	-0.067	-0.104	-0.075
	(-3.09)	*(-2.80)*	*(-1.59)*
Adj R^2	1.1%	0.7%	0.3%

Appendix A10: Control for the Extreme Inventory

This table examines the return predictability of relative basis after controlling for the extreme inventory. The Fama-Macbeth regression model is as follows:

$$ret_{i,t+1} = a_t + b_{1,t}RelatBasis_{i,t} + b_{2,t}Dummy\_HighInv_{i,t}$$
$$+b_{3,t}RelatBasis_{i,t} * Dummy\_HighInv_{i,t} + \lambda_t controls_{i,t} + \varepsilon_{i,t}$$

where ret_{t+1} is the commodity cumulative excess returns in month $t + 1$. We define *Dummy_ HighInv*=1 if the inventory of each commodity exceeds the 90 percentile of its time-series distribution, otherwise *Dummy_ HighInv* =0. Control variables include momentum and BasisMom. The sample is from January 1993 to June 2018. The *t*-statistics are reported in italic fonts below the corresponding coefficients, with the Newey-West adjustment with 12 lags.

	1	2	3	4
$RelatBasis_{i,t}$	0.022	0.021	0.020	0.022
	(3.21)	*(2.47)*	*(2.41)*	*(1.98)*
$Dummy\_HighInv_{i,t}$	-0.027	-0.024	-0.023	-0.023
	(-1.46)	*(-1.37)*	*(-1.21)*	*(-1.21)*
$RelatBasis_{i,t}$ * $Dummy\_HighInv_{i,t}$	-0.002	-0.011	0.003	0.003
	(-0.09)	*(-0.38)*	*(0.10)*	*(0.10)*
$ResidBasis_{i,t}$			-0.005	
			(-0.48)	
$TradtBasis_{i,t}$				-0.005
				(-0.48)
$Momentum_{i,t}$		0.012	0.011	0.011
		(2.18)	*(1.97)*	*(1.97)*
$BasisMom_{i,t}$		0.062	0.077	0.077
		(2.70)	*(2.97)*	*(2.97)*
Adj R^2	3.6%	10.7%	13.9%	13.9%