A Comprehensive Evaluation of Commodity Tracking Divergence

by

Colburn Hassman, Olga Isengildina-Massa, and Shamar Stewart

Suggested citation format:
Hassman, C., O. Isengildina-Massa, and S. Stewar. 2021. “A Comprehensive Evaluation of Commodity Tracking Divergence.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. [http://www.farmdoc.illinois.edu/nccc134].

A Comprehensive Evaluation of Commodity Tracking Divergence

Colburn Hassman, Olga Isengildina-Massa, and Shamar Stewart[1]

Paper prepared for the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, 2021.

[1] Colburn Hassman is a Graduate Student, Olga Isengildina-Massa is a Professor, and Shamar Stewart is an Assistant Professor in the Department of Agricultural and Applied Economics at Virginia Tech.

A Comprehensive Examination of Commodity ETF Tracking Divergence

Abstract

This paper investigates differences in returns between the ETF price, Net Asset Value, and Benchmark Asset Baskets for five popular futures-backed ETFs. We decompose tracking difference to examine the relative size of tracking differences attributable to managers versus the arbitrage process. Tracking differences attributable to managers is found to be significantly smaller than that attributable to the arbitrage process. We then test for average Tracking Differences using the Mincer-Zarnowitz Equation. We find evidence of bias in returns for multiple ETFs and demonstrate the usefulness of the decomposition. Furthermore, we investigate the dynamics of Tracking Error using a GARCH methodology. We find support that the volatility of the ETF effects Tracking Error but find no evidence that rolling futures contracts influences Tracking Error.

Keywords: Commodity ETFs, Tracking Errors, Futures-backed ETFs, ETF Performance

1 Introduction

Exchange Traded Funds (ETFs) are an important and growing portion of global financial markets, often playing a significant role in portfolio construction for professional asset managers and retail investors alike. ETFs were first introduced in the U.S. with the Standard & Poor's 500 Depository Receipt (SPDR) in 1993 but their history can be traced to *portfolio* or *program* trading introduced in the 1970s (Gastineau, 2010). Since their introduction, ETFs have grown to hold 5.7 trillion dollars as of 2019, and are expected to reach 10 trillion by 2024 (IPE, 2019). ETFs are part of the broader Exchange Traded Product (ETP) universe which also includes Exchange Traded Notes (ETNs) and Exchange Traded Commodities (ETCs). Similar to mutual funds, ETPs help investors construct diverse and sophisticated portfolios in an inexpensive and efficient manner. Unlike mutual funds, ETPs are traded throughout the day, increasing price transparency and allowing for intraday liquidity. Broad-index equity ETPs, such as the SPDR, allow investors to gain passive exposure to a large market-capitalization weighted basket of stocks across major sectors without having to rebalance the index, consider the inclusion of companies into the index, or manage dividend payments. Lastly, the ease of use and tax efficiency of ETPs make them increasingly popular tools for investors.

In the commodity space, ETPs provide unique benefits to investors. Buoyed by a low historical correlation with traditional investment markets (such as equities and bonds), commodities have become increasingly financialized and have grown in popularity amongst investors since the early 2000s. This low correlation augurs well for the diversification of investment portfolios (Gorton and Rouwenhorst, 2004). Around the same time, commodity markets saw a large increase in transaction volume and the transformation from primarily in-person pit trading to electronic trading (Irwin and Sanders, 2012). Despite the ease of access provided by electronic markets, gaining commodity exposure through the traditional futures markets continues to have several considerable drawbacks for retail investors. The first drawback is the relatively large size of futures contracts. For example, the size of a corn contract traded on the CME exchange is 5,000 bushels. At $4.75 a bushel, the notional value of one contract is $23,750, which may be too large for a standard investor to use as a diversifying portion of his/her portfolio.[1] The advent mini and micro contracts in recent years has partially ameliorated this issue but other problems persist. Unpredictable cash flows in the form of margin calls may cause liquidity issues for small investors. Investors may need to quickly deposit cash to cover losses in their future's position. Finally, because futures contracts expire, investors need to roll the contract forward to continuing gaining exposure. Commodity ETPs help address these challenges and provide added opportunities to invest in broad indexes or sub-indexes which include multiple commodities. The flexibility and the ease of use of commodity ETPs have undoubtedly increased their popularity, especially among retail investors.

Because the primary purpose of ETPs is to provide investors with exposure to underlying assets or strategies, the ability of ETPs to accurately track the returns of their benchmark should be considered the primary measure of success. In reality, differences in returns between the ETF and benchmark exist and at times are substantial. This difference in performance is a key concern for ETF investors who construct portfolios based of the stated benchmark but trade at the ETF price. Consistent bias in returns between the ETF and the benchmark implies that over the life of an investment, the two prices may drift apart considerably. Short-term tracking divergence may also have implications for a portfolio as diversification measures may fail in times of stress. It is thus important for investors to understand the sources and nature of tracking divergences when investing in ETPs.

The goal of this paper is to investigate tracking differences in five popular futures-backed commodity ETFs: three agricultural ETFs from Teucrium funds: CORN (Corn), SOYB (Soybean), and WEAT (Wheat), as well as two energy ETFs issued by USCF: USO (WTI Crude Oil) and UGA (US Gasoline). Futures-backed commodity ETFs are an under-studied portion of the ETF universe and face unique issues compared to ETFs employing a different replication method. Our first research question concerns the relative success of the ETF Manager and the ETF Creation and Redemption process in tracking the underlying asset. From where does Tracking Differences originate? The second research area is investigating average Tracking Differences, which are especially important for long-term investors. Are there long-term differences in returns, and from where do they originate? The final research question concerns the dynamics of the volatility of Tracking Differences. How does the volatility change over time and what factors effect it? We are able to provide additional insights into the nature and causes of tracking differences by reconstructing benchmark asset baskets where none are readily available. The findings of our study are useful to ETP investors and financial professionals who wish to better understand the risks associated with investing in futures-based commodity ETFs.

[1] $4.75 is the price of the December contract at the time of writing.

2 Data

The data from this paper comes from a variety of sources, including the Fund Manager, the Fund Prospectus, and the Bloomberg and Thomson Reuters Data services. We begin by analyzing the important attributes of the ETFs studied in this paper. Table 1 summarizes the relevant characteristics of each ETF as of March 30, 2021. As shown, the Agricultural ETFs (CORN, SOYB, and WEAT) hold multiple contract months in different proportions, while the two Energy commodities only hold a single contract month during the period of study. It is important to emphasize that these ETFs mimic the exposure to futures contracts, rather than the cash market. By holding futures contracts, the ETFs are not meant to track the spot price of the commodity. Irwin et al. (2020) and Guedj, Li, and McCann (2011) document large discrepancies between the performance of futures-backed commodity ETFs and commodity spot prices, often a source of confusion among novice commodity ETF investors.

The Expense Ratios, which are defined as "the amount of income required for the redemption value at the end of one year to equal the selling price of the Share", vary across ETF (Teucrium, 2020; USCF, 2020). The Expense Ratio is the amount of money the benchmark would need to appreciate by in order to cover the fees and operating expenses of the manager. This is an all-encompassing measure of Management Fees taking into account all charges and expenses incurred by the ETF manager. The agricultural ETFs all charge significantly higher expense ratios (2.47%-3.14%) compared to the energy ETFs (0.73% and 0.75%).

It is also interesting to note the relative size of ETF Assets Under Management (AUM). AUM refers to the value of all of the assets which back the ETF: the value of the ETF's portfolio. AUM can be thought of a measure of popularity as it shows the value invested in the ETF. The AUM of USO (around 3 billion USD) is twenty times the size of CORN (157 million USD), the next largest ETF in terms of AUM.

For each ETF, there are three primary measures used to analyze it's tracking performance. The first is daily open, high, low, and close (OHLC) prices which are collected from Bloomberg and Thomson Reuters databases. OHLC prices are a reflection of the price of the ETF traded on the exchange, the actual prices at which investors bought and sold the ETF.

The second is Net Asset Value (NAV). The NAV is the ETF manager's portfolio assets minus its liabilities. In ETFs, the value is often reported on a per-share basis by dividing this value by the total number of shares outstanding. The NAV thus represents the value of the portion of the manager's portfolio which "backs" each ETF share. This information is readily available to investors, often through the ETF Manager's website or public sources. Our data is collected through the Bloomberg data service.

Table 1: ETF Information

	CORN	**SOYB**	**WEAT**	**USO**	**UGA**
ETF Exchange	NYSE Arca	NYSE Arca	NYSE Arca	NYSE Arca	NYSE Arca
Futures Contracts	CBOT Corn	CBOT Soybean	CBOT Wheat	NYMEX WTI Crude Oil	RBOB Gasoline
Holdings	2nd to Expire (35%) 3rd to Expire (30%) Following December (35%)	2nd to Expire (35%) 3rd to Expire (30%) Following November (35%)	2nd to Expire (35%) 3rd to Expire (30%) Following December (35%)	Next to Expire*	Next to Expire
ETF Price	$16.28	$20.49	$5.82	$41.17	$31.95
ETF Assets Under Management	$157M	$87M	$78M	$3,080M	$122M
Expense Ratio	2.47%	2.50%	3.14%	0.73%	0.75%
ETF Manager	Teucrium Trading	Teucrium Trading	Teucrium Trading	USCF	USCF

Data collected from ETF Manager Website, ETF Prospectus, and Thomson Reuters data service

The third measure of ETF performance is its Benchmark. This is the "goal" portfolio which the ETF managers strive to recreate. Benchmark prices reflect the value of the asset basket which the ETF is meant to track, according to the ETF investment goals. In our case, these are the settlement prices of the included futures contracts described in Table 1. While many studies which investigate Tracking Differences have benchmark values readily available (such as the SP 500 Index or Bloomberg Commodity Index), there are no such benchmark values available for the ETFs investigated in this paper. In order to reconstruct the benchmark asset baskets for each ETF, daily closing futures prices for all of the included contracts were collected from Bloomberg, Thomson Reuters, and Quandl. Where ETFs contained multiple futures contracts, as in the case of CORN, SOYB, and WEAT, the asset basket was recreated using the target weightings found in the fund prospectus:

$$B_t = \sum_{i=1}^{N} P_t^F W_F \tag{1}$$

where B is the benchmark value, N is the number of futures contracts included in the benchmark, P_t^F is the price of futures contract F on day t and W_F is the target weighting of futures contract F. Notice that there is no accounting for management fees, transaction costs, or other limitations of the manager in the benchmark.

One of the challenges for long-term passive commodity investors utilizing futures contracts to gain exposure is that contracts expire regularly. With commodity-backed ETFs, the issue of expiring contracts is handled by the ETF manager rather than the ETF investor. The futures-based ETF Managers approach this problem by *rolling* forward futures contracts on a set, predetermined schedule. This roll period can last from one to multiple days during which time the fund sells the contracts closer to expiration and buys the contracts farther from expiration. The roll period timing and procedures for each ETF are described in it's prospectus and the exact roll dates are collected from the fund managers. These roll dates are vital to properly reconstructing the benchmark as they mark a change of the contract months included.

While some indexes, such as the Bloomberg Commodity Index, provide guidance as to the portion of contracts which are rolled forward on each day of a multi-day roll period, the ETF managers do not provide such information (Bloomberg, 2016). This leads to uncertainty regarding the exact composition of the benchmark during roll periods. Because of this uncertainty, in order to not artificially induce tracking differences, we exclude roll days from our initial analysis of the ETF benchmark. However, we assume that by the end of the last day of the roll period the fund manager has completed their transactions and the ETF holds the new contracts.

Days when either the major stock exchanges or the relevant CME Group exchanges were closed are also excluded from the analysis. Holidays and market closures are largely synchronous between the major US markets, but not always. An example of a conflicting case was December 5, 2018, a National Day of Mourning for former President George H.W. Bush. On this day, the major stock markets (on which ETFs trade) were closed, as well as the CME Group Equity and Interest Rate

Table 2: Price Summary Statistics

		CORN	SOYB	WEAT	USO	UGA
Price	Min	11.67	13.34	4.86	8.24	8.90
	Median	21.24	18.97	8.94	12.78	32.63
	Mean	24.82	19.66	11.03	17.63	38.61
	Max	52.67	28.85	25.30	39.36	65.71
	Std Dev	10.19	4.37	5.78	9.83	14.11
NAV	Min	11.70	13.38	4.89	8.14	8.52
	Median	21.25	18.98	8.94	12.80	32.61
	Mean	24.82	19.66	11.03	17.64	38.62
	Max	52.68	28.77	25.17	39.48	65.44
	Std Dev	10.19	3.48	5.77	9.84	14.13
Bench-mark	Min	320.48	834.26	421.88	28.35	0.49
	Median	394.30	990.18	535.71	56.43	1.84
	Mean	437.99	1,057.96	574.55	62.87	2.02
	Max	774.23	1,632.46	916.03	110.53	3.40
	Std Dev	99.28	179.39	112.09	20.33	0.64
	N	2,031	2,079	2,040	1,334	2,039

products markets. CME commodity markets remained open. Including these dates in the dataset would create artificial tracking error between the ETF benchmark (which changed) and the ETF price (which did not.)

USO is a special case wherein the asset basket holding criteria mid-2013 and again in early 2020 around the COVID-19 pandemic-induced energy market volatility. Because of this change, analysis including the USO benchmark is evaluated beginning July 2013 and ending January 2020. Table 2 displays summary statistics for ETF price, NAV, and benchmark for each ETF. All ETF prices and NAVs are reported in dollars per share. Benchmark values for USO are reported in dollars per barrel, and dollars per gallon in the case of UGA. Agricultural benchmark values are reported in cents per bushel.

Figure 1 shows changes in ETF price, NAV, and benchmark during the period of study. Our research focuses on differences between returns. We construct log returns as show in Equation 2 for ETF Price, NAV, and Benchmark.

$$R_d = ln(\frac{P_t}{P_{t-1}}) \cdot 100 \tag{2}$$

Table 3 includes summary statistics for returns of each measure while Figure 2 shows these values over time. The COVID-19 Pandemic's effect on markets are clearly visible in UGA but not so in CORN, SOYB, and WEAT (recall that this period is excluded from the current analysis of USO). For each measure of each ETF, the mean return was negative over the sample period. The two energy ETFs, USO and UGA, are more volatile then their agricultural counter parts, both in terms of the standard deviation of returns and the minimum and maximum returns.

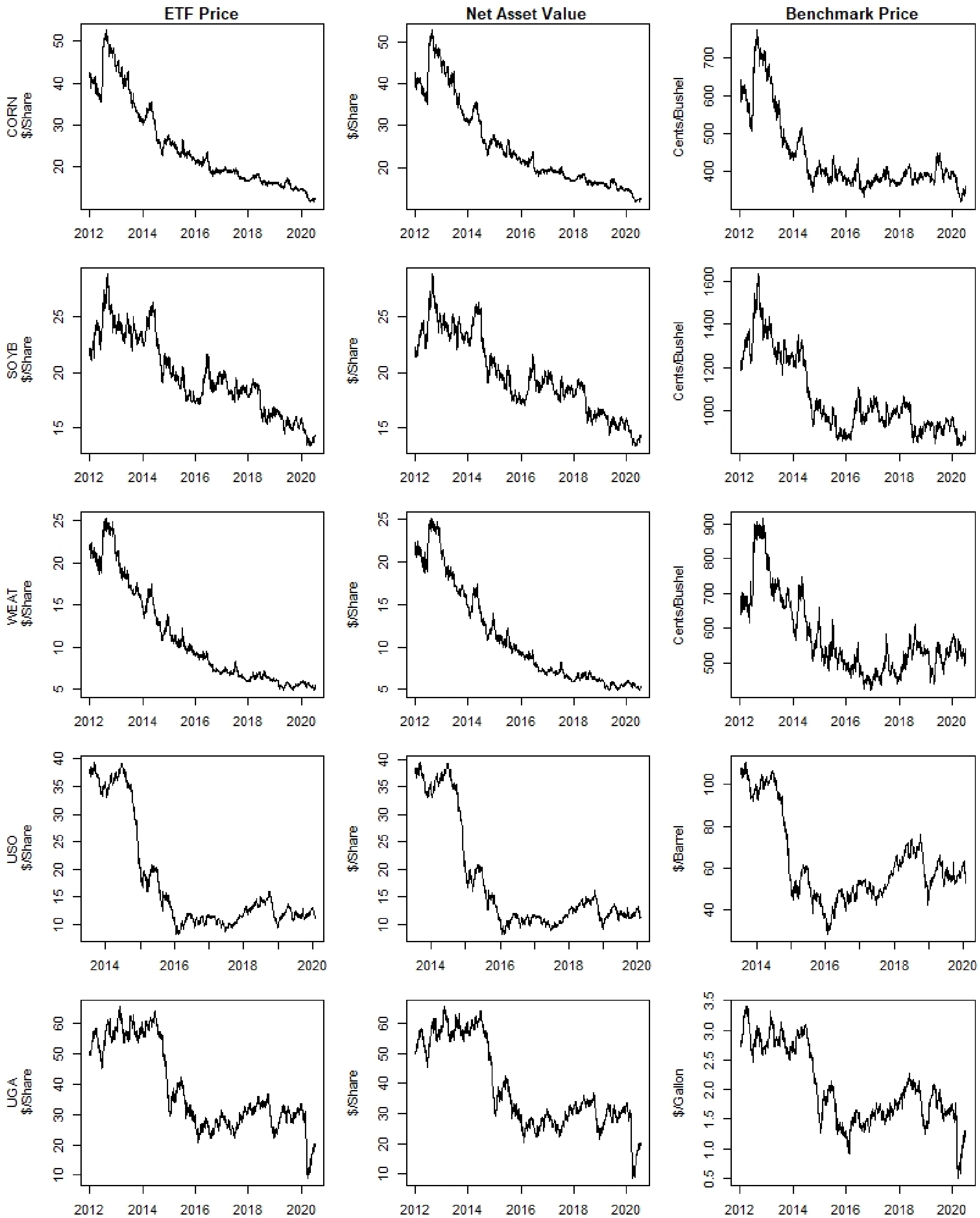

Figure 1: ETF Price, NAV, and Benchmark over time

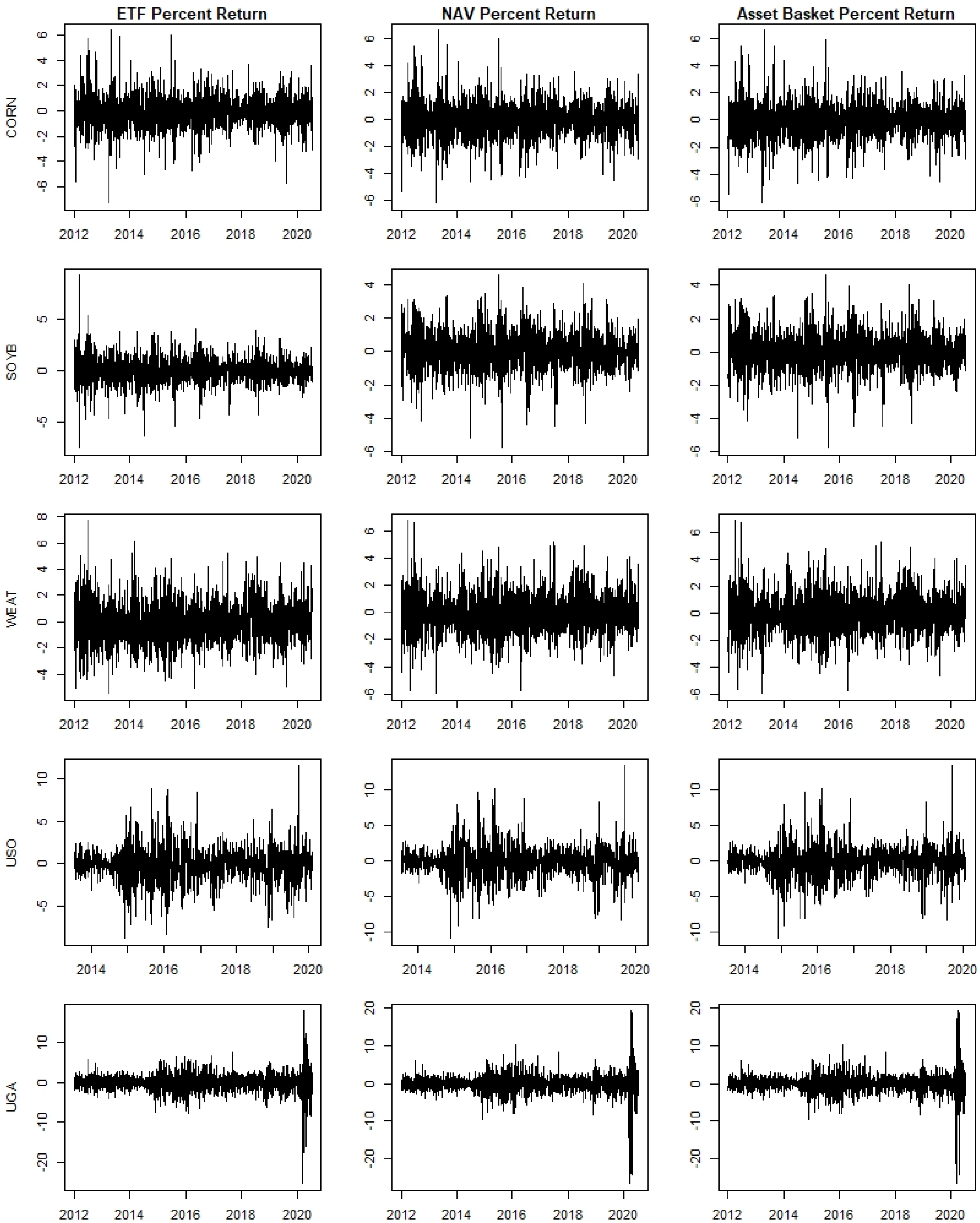

Figure 2: ETF, NAV, and Benchmark Returns over time

Table 3: Return Summary Statistics

		CORN	SOYB	WEAT	USO	UGA
Price	Min	-7.26	-7.55	-5.45	-8.68	-25.29
	Median	-0.05	0.00	-0.14	0.03	-0.02
	Mean	-0.06	-0.02	-0.07	0.05	-0.06
	Max	6.34	9.22	7.72	11.47	17.95
	Std Dev	1.22	1.15	1.50	2.03	2.27
NAV	Min	-6.19	-5.77	-5.96	-10.80	-26.52
	Median	-0.05	0.00	-0.12	0.02	0.01
	Mean	-0.06	-0.02	-0.07	-0.05	-0.06
	Max	6.59	4.54	6.82	13.43	19.30
	Std Dev	1.22	1.03	1.43	2.16	2.45
Bench-mark	Min	-6.10	-5.77	-5.91	-10.79	-26.50
	Median	-0.04	0.01	-0.11	0.02	0.02
	Mean	-0.04	0.00	-0.05	-0.05	-0.06
	Max	6.59	4.59	6.85	13.42	19.29
	Std Dev	1.22	1.03	1.42	2.16	2.45
	N	2,030	2,078	2,039	1,333	2,038

3 Tracking Differences

Johnson et al. (2013) argue that there has been considerable heterogeneity in defining measures of ETP tracking performance both in industry and in academic research. In an attempt to clarify this matter for reporting purposes, the European Securities and Markets Authority (ESMA) issued guidelines defining both Tracking Difference and Tracking Error, two common words often used incorrectly and interchangeably. *Tracking Difference* is the difference in returns between the ETP and the benchmark index. *Tracking Error* is the volatility of the difference between the return of the ETP and the benchmark index (ESMA, 2012). Tracking Differences between ETF prices P_E and benchmark prices P_B daily (d) can thus be computed by subtracting one return from the other, as shown below.

$$TD_t = R_{E,t} - R_{B,t} = ln\left(\frac{P_{E,t}}{P_{E,t-1}}\right) \cdot 100 - ln\left(\frac{P_{B,t}}{P_{B,t-1}}\right) \cdot 100 \tag{3}$$

As stated previously, this definition of Tracking Difference is very meaningful for investors, who transact at ETF prices but presumably build portfolios based off the stated benchmark of the fund. Short-term investors may be bewildered and frustrated when the returns of the ETF and the benchmark do not align. Investors with longer holding periods are likely especially concerned with average differences of returns, which may compound significantly over the extended holding period.

So far, we have only discussed two of the three measures as it relates to Tracking Differences: the ETF Price and Benchmark. By including the NAV, we are able decompose this Total Tracking Difference into two separate components, show in Figure 3. Tracking Differences which lead to valuable insights as to the nature and causes of this issue (Elton et al., 2002; Aber et al., 2009). The first component of Total Tracking Difference is Managerial Tracking Difference (TD_M): Tracking Differences attributable to the ETF manager. This is calculated as the difference between NAV and Benchmark returns. The stated investment goal of all ETFs studied in this paper is that the NAV, rather than the ETF Price, replicates the exposure of the benchmark (USCF, 2020; Teucrium, 2020) This is a subtle but important distinction. By holding a portfolio which replicates the benchmark, ETF managers only control the differences between the benchmark and the NAV. The benchmark can be thought off the goal portfolio while the NAV is the value of the actual fund portfolio, inclusive of fees. Frino and Gallagher (2001) argued that tracking divergence is unavoidable due to market friction, as the benchmark index is calculated as if transactions occur instantaneously and without costs. These costs would be reflected in NAV performance. Other potential causes of TD_M include unintentional deviations from the benchmark, including cash drag and weightings drift (Gastineau, 2010)[2].

The second component of Total Tracking Difference concerns the ability of the ETF price to reflect the NAV, the value of the ETF manager's portfolio. Arbitrage Tracking Difference (TD_A) is the difference between ETF price and the Net Asset Value. Gallagher and Segara (2006) noted that the price of an ETP is determined by the supply and demand characteristics for the ETP itself. These characteristics might be misaligned with those of the underlying asset leading to a misalignment of returns and exposure. Authorized Participants[3] of the fund can exchange the underlying assets of the ETF for shares of the ETP via the Creation and Redemption (Arbitrage) process, keeping the two prices in line. The exact process for creating and redeeming ETF shares varies by ETF but it is far from instantaneous and frictionless for the ETFs we study.[4] Hill, Nadig, and Hougan (2015) pointed out that the arbitrage gap (the price at which it makes sense for ETF Authorized Participants to step in) varies with the liquidity of the underlying securities and related costs. In some ETFs, the gap can be as small as 1 cent, and substantially larger in others.

To summarize, the ETP's success in tracking the underlying benchmark will depend both on the skill of the manager to keep the NAV in line with the the benchmark and the ease of the Creation and Redemption process to keep the ETF price aligned with NAV. The sum of Managerial and Arbitrage Tracking Differences forms Total Tracking Differences.

$$TD_t = TD_{A,t} + TD_{M,t} = (R_{E,t} - R_{N,t}) + (R_{N,t} - R_{B,t})$$

Table 4 provides summary statistics of each of the three tracking differences for each ETF while Figure 4 displays the tracking difference over time. We find that in absolute terms, Arbitrage

[2]Cash drag refers to ETF managers holding more cash then necessary, thus weighing down returns. Though mostly discussed in the context of equity ETFs, cash drag is especially important for future-backed commodity ETF managers who buy futures on margin and thus have a significant cash position. Weightings Drift is the idea that as the contents of the ETF manager's portfolio do not move in lock-step, the initial weightings of each asset will drift from the initial weightings, requiring portfolio rebalancing.

[3]Authorized Participants are traditionally large financial institutions, such as major banks, who fulfill the role of arbitrager between the ETF price and the NAV.

[4]The user is directed to the appendix for a diagrammatic explanation of this process.

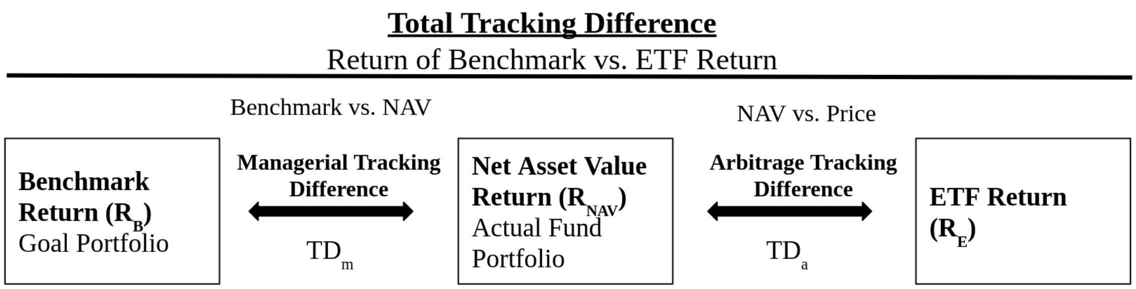

Figure 3: Tracking Difference Decomposition

Table 4: Tracking Difference Summary Statistics

	Value	CORN	SOYB	WEAT	USO	UGA
TD	Min	-1.33	-7.95	-4.25	-6.01	-7.76
	Median	-0.02	-0.02	-0.01	-0.01	1.01E-04
	Mean	-0.01	-0.02	-0.02	9.70E-04	2.88E-03
	Max	2.03	8.97	4.18	3.98	8.2
	Std Dev	0.25	0.52	0.52	0.57	0.64
TD_M	Min	-0.16	-0.21	-0.23	-0.01	-0.02
	Median	-0.01	-0.01	-0.01	-5.34E-04	-1.27E-03
	Mean	-0.01	-0.02	-0.02	8.55E-04	-4.38E-04
	Max	0.28	0.11	0.1	0.02	0.04
	Std Dev	0.02	0.02	0.02	4.08E-03	4.38E-03
TD_A	Min	-1.33	-7.9	-4.21	-6	-7.78
	Median	-2.23E-03	-4.42E-03	6.57E-04	-0.01	1.74E-03
	Mean	-1.56E-04	-1.41E-03	-9.76E-04	1.16E-04	3.32E-03
	Max	1.75	9	4.23	3.96	8.2
	Std Dev	0.25	0.52	0.52	0.57	0.64
	N	2030	2078	2039	1333	2038

Summary information for Total Tracking Difference, Managerial Tracking Difference, and Arbitrage Tracking Difference.

Tracking Differences are much larger than Managerial Tracking Differences. This implies that the ETF managers do a relatively good job at keeping their portfolios aligned with the benchmark and the arbitrage process is relatively less successful in keeping the ETF price aligned with NAV. This finding conflicts with similar studies in other ETFs, namely Elton et al. (2002) which investigates Standard and Poor's Depository Receipts (SPDRs) and Gallagher and Segara (2006) which studies ETFs traded on the Australian Stock Exchange. These studies find that the Tracking Differences attributable to the Manager are larger than those attributable to the Arbitrage process. These conflicting results across ETFs highlights the need to consider specific characteristics of each futures-backed ETFs individually.

TD_A is similar to the concept of ETF premium and discounts to NAV, the percent above or below the price of price of the ETF is compared to NAV, explored in previous studies and show in Equation 5. While the calculation differs, both metrics capture the differences between NAV

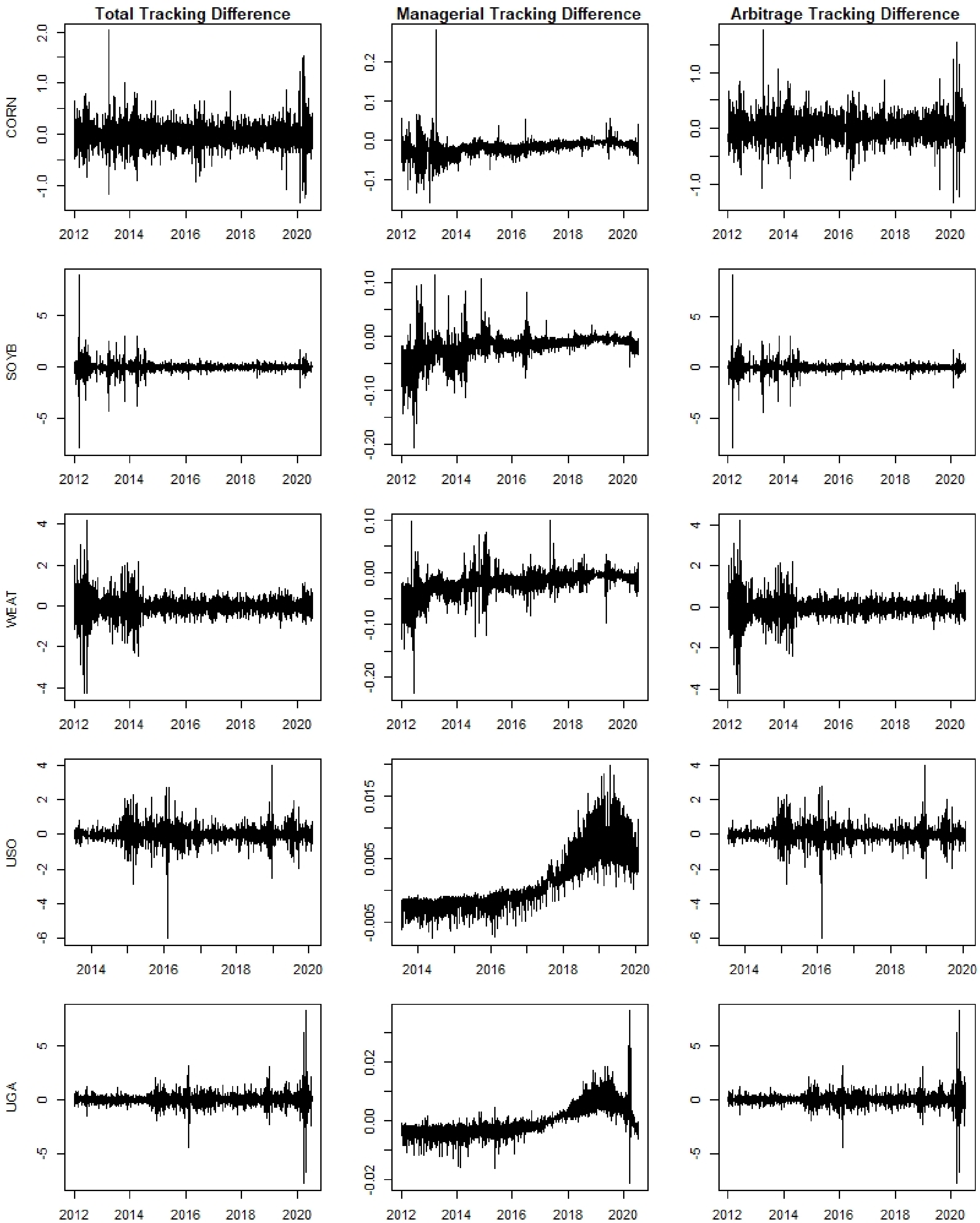

Figure 4: TD, TD_M, and TD_A

and the ETF Price, effectively analyzing the effectiveness of the Creation and Redemption Process. The research is divided as to the relative importance of premium and discounts to NAV (Engle and Sarkar, 2006; Aber et al., 2009), again highlighting the need to analyze ETFs individually an avoid extending findings inappropriately.

$$Premium_t = \frac{(P_{E,t} - P_{N,t})}{P_{N,t}} \tag{5}$$

Because TD_M is so small in the ETFs studied, we find the size of TD_A to be similar to the size of TD. As previously mentioned, other research in equity ETFs found TD_M to be larger than TD_A. In studies of commodity ETFs, especially futures-backed commodity ETFs, the focus has largely been on TD_A. This is likely because there is often no benchmark readily available and TD_A does not include the benchmark when calculating. Given the conflicting results regarding the relative size of TD_A and TD_M, it is important that ETF investors recognize that differences between ETF price and NAV are not the only source of Tracking Differences.

4 Average Tracking Differences

Average Tracking Differences refers to long-term differences between two returns. For a long-term investors, average tracking differences are extremely important. A small average daily tracking differences compounded over a significantly long holding period may lead to substantial differences in total returns.

In order to investigate average tracking differences, we utilize the methodology of Mincer and Zarnowitz (1969), widely used throughout the literature wherein the two returns are regressed on one another to judge the accuracy and bias. This analysis of Tracking Divergences differs from the previously defined equations but leads to meaningful insights as to the nature of TD.

We begin by testing the stationarity of returns, a necessary condition for meaningful regression (Dickey and Fuller, 1979). Table 5 reports the results of the Augmented Dickey-Fuller Test for Unit Root with associated t-statistics and p-values. The null hypothesis of unit root is rejected in each case, indicating stationarity.

We define three Mincer-Zarnowitz equations corresponding to TD, TD_M and TD_A:

$$\text{Total Tracking Difference:} R_t^E = \alpha + \beta R_t^B + \epsilon_t^{TD} \tag{6}$$

$$\text{Managerial Tracking Difference:} R_t^N = \alpha + \beta R_t^E + \epsilon_t^{TD_M} \tag{7}$$

$$\text{Arbitrage Tracking Difference:} R_t^E = \alpha + \beta R_t^N + \epsilon_t^{TD_A} \tag{8}$$

where α is a measure of systematic bias and β is a measure of risk. Bias is how much on average the dependent returns is above or below the explanatory return. An unbiased ETF will show an

Table 5: Augmented Dickey Fuller Test on Returns

		Statistic	p_value
ETF	CORN	-9.292	< 0.01
	SOYB	-10.399	< 0.01
	WEAT	-10.189	< 0.01
	USO	-7.801	< 0.01
	UGA	-9.331	< 0.01
NAV	CORN	-9.303	< 0.01
	SOYB	-10.219	< 0.01
	WEAT	-10.227	< 0.01
	USO	-7.903	< 0.01
	UGA	-9.328	< 0.01
Benchmark	CORN	-9.326	< 0.01
	SOYB	-10.173	< 0.01
	WEAT	-10.202	< 0.01
	USO	-7.903	< 0.01
	UGA	-9.327	< 0.01

$\alpha = 0$. A negative (positive) α coefficient indicates that the daily ETF return is smaller (larger) than the benchmark return. The coefficient estimates are interpreted as average daily differences in percent returns.

Risk, reflecting in β, is a measure of whether the volatility of the base return is being properly transferred to the dependent return. β is an elasticity, a measure of the change in y brought about by x. The interpretation of β in this context is analogous the interpretation of the same coefficient in the Capital Asset Pricing Model (CAPM). An ETF with perfect unity has a $\beta = 1$. If β is greater (smaller) than one, the volatility of the dependent return is on average larger (smaller) than that of explanatory return. Both α and β capture average attributes of tracking differences.

Table 6: Mincer-Zarnowitz Results

			Estimate	SE	t score	p value	R^2
TD	CORN	α	-0.0152	0.0055	-2.7802	0.0055	0.9592
		β	0.9814	0.0045	-4.1483	< 0.0001	
	SOYB	α	-0.0168	0.0113	-1.4874	0.1371	0.7974
		β	0.9910	0.0110	-0.8218	0.4117	
	WEAT	α	-0.0174	0.0114	-1.5222	0.1281	0.8814
		β	0.9883	0.0080	-1.4507	0.1472	
	USO	α	-0.0036	0.0146	-0.2465	0.8053	0.9310
		β	0.9047	0.0068	-14.1098	< 0.0001	
	UGA	α	-0.0038	0.0129	-0.2916	0.7706	0.9336
		β	0.8951	0.0053	-19.8330	< 0.0001	
TD_M	CORN	α	-0.0143	4.36E-04	-32.6770	< 0.0001	0.9997
		β	0.9999	3.58E-04	-0.3954	0.6929	
	SOYB	α	-0.0154	4.90E-04	-31.4524	< 0.0001	0.9995
		β	0.9954	4.75E-04	-9.6815	< 0.0001	
	WEAT	α	-0.0156	4.86E-04	-32.0783	< 0.0001	0.9998
		β	1.0038	3.42E-04	10.9975	< 0.0001	
	USO	α	0.0009	1.12E-04	7.6727	< 0.0001	> 0.9999
		β	1.0000	5.17E-05	0.8559	0.3922	
	UGA	α	-0.0004	9.71E-05	-4.5103	< 0.0001	> 0.9999
		β	1.0000	3.97E-05	0.1166	0.9073	
TD_A	CORN	α	-0.0012	0.0054	-0.2250	0.8220	0.9596
		β	0.9816	0.0045	-4.1174	< 0.0001	
	SOYB	α	-0.0015	0.0113	-0.1319	0.8951	0.7976
		β	0.9955	0.0110	-0.4114	0.6812	
	WEAT	α	-0.0020	0.0114	-0.1789	0.8581	0.8814
		β	0.9845	0.0080	-1.9371	0.0529	
	USO	α	-0.0044	0.0146	-0.2997	0.7645	0.9310
		β	0.9047	0.0067	-14.1170	< 0.0001	
	UGA	α	-0.0034	0.0129	-0.2613	0.7939	0.9336
		β	0.8951	0.0053	-19.8351	< 0.0001	

Results of Mincer-Zarnowitz Regression for Total Tracking Difference, Managerial Tracking Difference, and Arbitrage Tracking Difference.

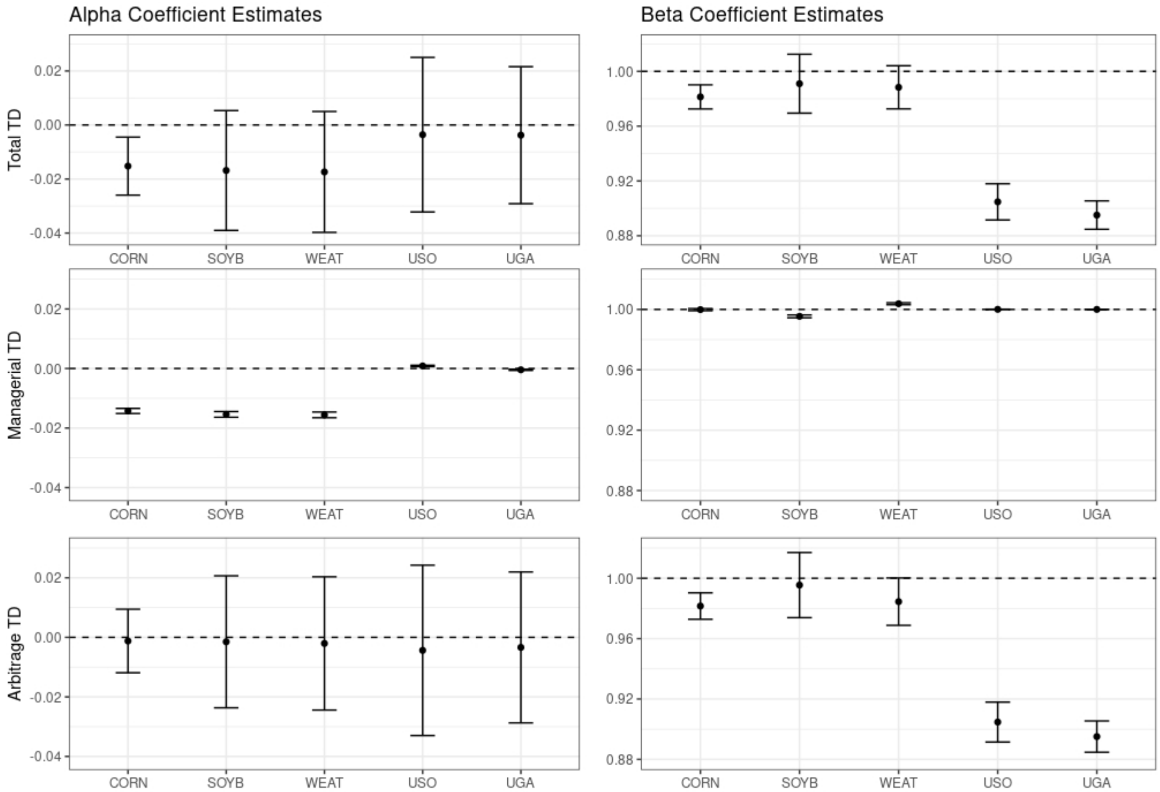

Figure 5: Mincer-Zarnowitz Coefficient Estimates with 95% Confidence Bars

Table 6 displays the results of linear regression for TD, TD_M, and TD_A. The significance of β is relative to 1, while all others are relative to 0. Figure 5 visualizes the coefficient estimates with 95% confidence bars. Out findings for each TD are discussed in the following sections.

4.1 Total Tracking Difference

The α coefficient estimates for Total TD are statistically difference from zero only in the case of CORN. All α estimates are below zero which indicates that on average, the ETF returns less than the benchmark for the ETFs studied. This phenomenon is likely caused by multiple factors including transaction costs, cash drag, and the expense ratio of the ETF. While the coefficient estimates may seem small, recall that these are differences in daily returns which compound over time. In the case of WEAT for example, the coefficient estimate of -0.0174% compounded over 252 trading days in a year implies an annual difference in ETF and Benchmark returns of -4.29%.

The β estimates for Total TD are significantly different from one in the case of CORN, USO, and UGA. For all of the ETFs studied, the β estimates are below one, indicating that on average, the volatility of the ETF is less than the volatility of the benchmark. This is especially true in the case of USO and UGA, where the ETF is approximately 10% less volatile than the Benchmark.

4.2 Managerial Tracking Difference

Decomposing TD into TD_M and TD_A is very informative for isolating the source of average tracking differences. It is clear that the negative α coefficient found in Total Tracking Differences is the result of the ETF Manager, especially for the agricultural ETFs. For TD_M, CORN, SOYB, and WEAT all the α coefficient estimates which are significant below zero, indicating that on average, the returns that the manager's achieve are less than the returns achieved by the benchmark. As discussed earlier, CORN, SOYB, and WEAT all have significantly higher expense ratios than USO or UGA which likely contributes to the differences in estimates.

Additionally, SOYB and WEAT have β estimates statistically different than one, though the estimates are very close to unity. The β coefficents less than one found in Total Tracking Differences are not explained by this compenent of Total Tracking Difference. In other words, the Manager's portfolio is just as volatile as the Benchmark. The high R^2 values indicate that variation in the NAV is almost perfectly explained by variation in the benchmark: Managers do a relatively good job of replicating Benchmark returns.

4.3 Arbitrage Tracking Difference

The α coefficient results for Arbitrage Tracking Difference indicates that on average, the return of the ETF are similar to the return of the NAV, but with larger uncertainty in coefficient estimates than Managerial Tracking Difference

In the case of CORN, USO, and UGA ETFs all have β estimates statistically different than one at the 5% level, with the USO and UGA estimates having similar magnitudes. WEAT is significant at the 10% level. This indicates that the variation in the returns of the ETF do not reflect the variation in the NAV, the ETF being less volatile. In other words, the failure of the ETF returns to match the volatility of the the Benchmark is due to the Arbitrage process, rather than the ETF Manager.

The R^2 values for the Arbitrage Tracking Difference equation differ across commodities, with CORN and SOYB having the best and worst fit, at 0.96 and 0.80 respectively. For both USO and UGA, approximately 93% of the variation of the ETF can be explained by the variation in the NAV. These values are significantly lower than the R^2 values of the Managerial Tracking Difference regression, which are all greater than 0.99, indicating larger tracking differences.

To summarize the results of the average tracking difference analysis, the additivity of TD_M and TD_A to form TD allows us to make judgements about the causes of average TD. Negative α estimates likely the result of managers, especially for the agricultural ETFs, the NAV returns less than the benchmark. This implies that over the investment period, the ETF will return less than the benchmark. β estimates less than one are likely the results of the arbitrage process rather than the manager: the volatility of the ETF does not reflect the price action of the NAV.

That CORN, SOYB, and WEAT estimates and USO and UGA estimates are grouped together is of note. Not only do they cover different groups of commodities (agricultural versus energy), but they also come from different fund managers (Teucrium and USCF). The finding highlights the heterogeneity of tracking differences across different ETPs, even amongst a small subset such as futures-backed commodity ETFs.

4.4 Tracking Error: the Volatility of Tracking Differences

In order to further analyze the relative size of TD_M and TD_A, and their contribution to TD, we turn to the concept of tracking error, the volatility of tracking differences. A simple summation of daily tracking differences would be insufficient as daily tracking differences with opposite signs (positive or negative) offset each other. While the total sum or average of daily tracking differences is informative regarding the bias of such differences, investors may also be concerned with the volatility of tracking differences, as there may be significant short-term implications for their portfolio versus the benchmark.

Without a common methodology defined by regulators, there are multiple calculations of tracking error. The three primary ways of measuring tracking error found in previous research are as the Average Absolute Difference in Returns (AARD), the Standard Deviation of Return Difference (SDRD), and as the standard error of the residuals of the Mincer-Zarnowitz equations defined above. The equations below show tracking error in terms of TD but can also be applied to TD_M and TD_A.

$$\text{Average Absolute Difference in Returns: } \frac{\sum_{t=1}^{N} \mid TD \mid}{N} \tag{9}$$

$$\text{Standard Deviation of Return Differences: } \sqrt{\frac{\sum_{t=1}^{N}(TD_t - \bar{TD})^2}{N-1}} \tag{10}$$

$$\text{Standard Error of the Residuals: } \sqrt{\frac{\sum_{t=1}^{N} \epsilon_t^2}{df}} \tag{11}$$

where ϵ are the residuals from the Mincer-Zarnowitz equation show in Equation 6 and df are the degrees of freedom from that same equation.

We follow similar notation for tracking error as tracking differences, wherein TE is the volatility of TD, TE_M is the volatility of TD_M, and TE_A is the volatility of TD_A. As show in Table 7 we find the size of TE_A to be significantly larger than TE_M by all three metrics in all the ETFs we study. This again indicates that the returns of ETF price and NAV differ more than the returns of NAV and the benchmark. This finding reiterates that the creation and redemption process does a relatively poor job of keeping the price of the ETF in line with the NAV compared to the manager's effort to align the NAV and benchmark.

Additionally of note are the differences between agricultural and energy ETF TE_M metrics. The agricultural ETFs have much larger TE_M versus energy ETFs. This implies that the energy ETF manager (USCF) does a better job of keeping NAV aligned with the benchmark over the sample period compared to the agriculture ETF manager (Teucrium Funds). These differences in magnitude may be explained by the energy ETFs holding only one futures contract (rather than three in the case of CORN, SOYB, and WEAT) or the lower managerial fee.

Table 7: Metrics of Tracking Error

	ETF	TD	TDm	TDa
AARD	CORN	0.1809	0.0167	0.1805
	SOYB	0.2646	0.0181	0.2640
	WEAT	0.3352	0.0191	0.3347
	USO	0.3765	0.0030	0.3765
	UGA	0.3800	0.0032	0.3800
SDRD	CORN	0.2475	0.0196	0.2461
	SOYB	0.5159	0.0228	0.5156
	WEAT	0.5152	0.0226	0.5156
	USO	0.5710	0.0041	0.5710
	UGA	0.6382	0.0044	0.6382
SER	CORN	0.2465	0.0196	0.2452
	SOYB	0.5160	0.0223	0.5157
	WEAT	0.5151	0.0219	0.5152
	USO	0.5328	0.0041	0.5327
	UGA	0.5839	0.0044	0.5838

AARD: Average Absolute Difference in Returns

SDRD: Standard Deviation of Return Differences

SER: Standard Error of the Residuals.

Between tracking difference and tracking error, which measurement is most meaningful for investors? Gastineau (2010) argues that average tracking difference should be the preferred framework for assessing fund manager performance, writing that "the fund manager's objective should be to achieve the best possible performance for investors, not the smallest possible tracking error" (page 162). Johnson et al. (2013) also argue that for long-term long-only investors, average tracking difference is the more appropriate measure but point out that investors who have a mandate to closely track an index, who short sell the ETF to express a market opinion or to hedge exposure, or otherwise use the ETF for hedging or risk management purposes may find tracking error to be a more valuable metric. For all of the ETFs we investigate, the stated goal of the fund manager is to minimize tracking error between the return of the ETF's NAV relative to the benchmark, rather than achieve positive returns relative to the benchmark (positive tracking difference).

Our final research question concerns the dynamics of Tracking Error, the volatility of tracking differences. Due to the dominance of TE_A over TE_M, we focus our attention on TD_A. Because TD_A is a function of only the ETF price and NAV, and not the benchmark, we are able to expand our dataset by including previously excluded roll dates and test for the effect of *rolling* on these dates. Additionally, because the benchmark is not included in the analysis, USO is now evaluated for the same time period as the other ETFs (January 2012 to July 2020). This allows us to capture USO TD_A during the volatility brought about by the COVID-19 pandemic. Updated summary

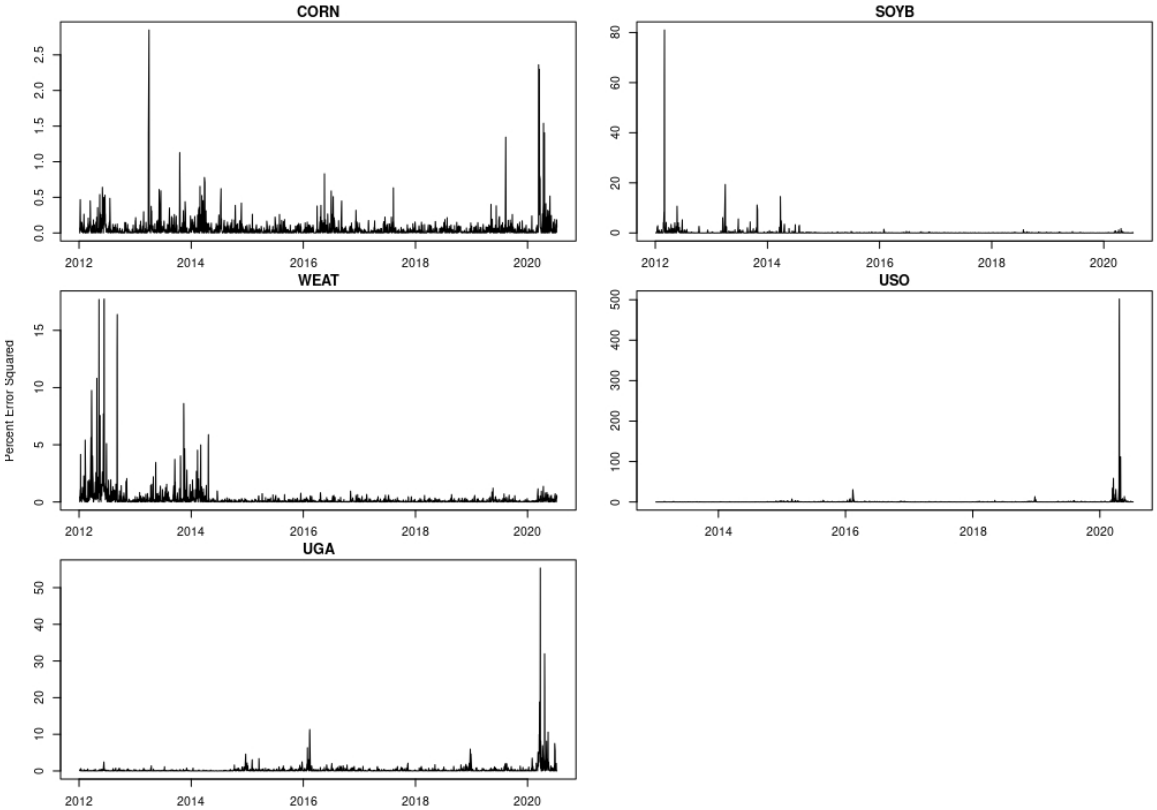

Figure 6: Squared Non-systematic TD_A

statistics can be found in the Appendix.

5 Modeling Tracking Error

Several previous studies have identified tracking error as being non-constant (Johnson et al. (2013); Perera, Bialkowski, and Bohl (2018); Qadan and Yagil (2012)). They find that the volatility of Tracking Differences changes over time, with periods of relatively high and low variance. This has implications for investors, as tracking performance changes over time. We begin by examining the residuals of the Arbitrage Mincer-Zarnowitz equation, Equation 8. Figure 6 shows the squared residuals. The variation in the residuals appears to change over time, with periods of high and low variation grouped together. Figure 7 shows the ACF plots for squared non-systematic TD_A. Autocorrelation in squared residuals indicates that there is a relationship in volatility from one time period to the next. In each ETF, there is is significant autocorrelation in the squared residuals for multiple lag periods. Together, these visualizations suggest that the variance of the tracking differences, TE, studied is likely non-constant overtime. We formality test this by utilizing the Ljung-Box Procedure on 20 lag periods, the results of which are presented in Table 8 (Ljung and Box, 1978). For each ETF, we find evidence of autocorrelation in the squared residuals, further supporting the presence of heteroskedaticy of residuals.

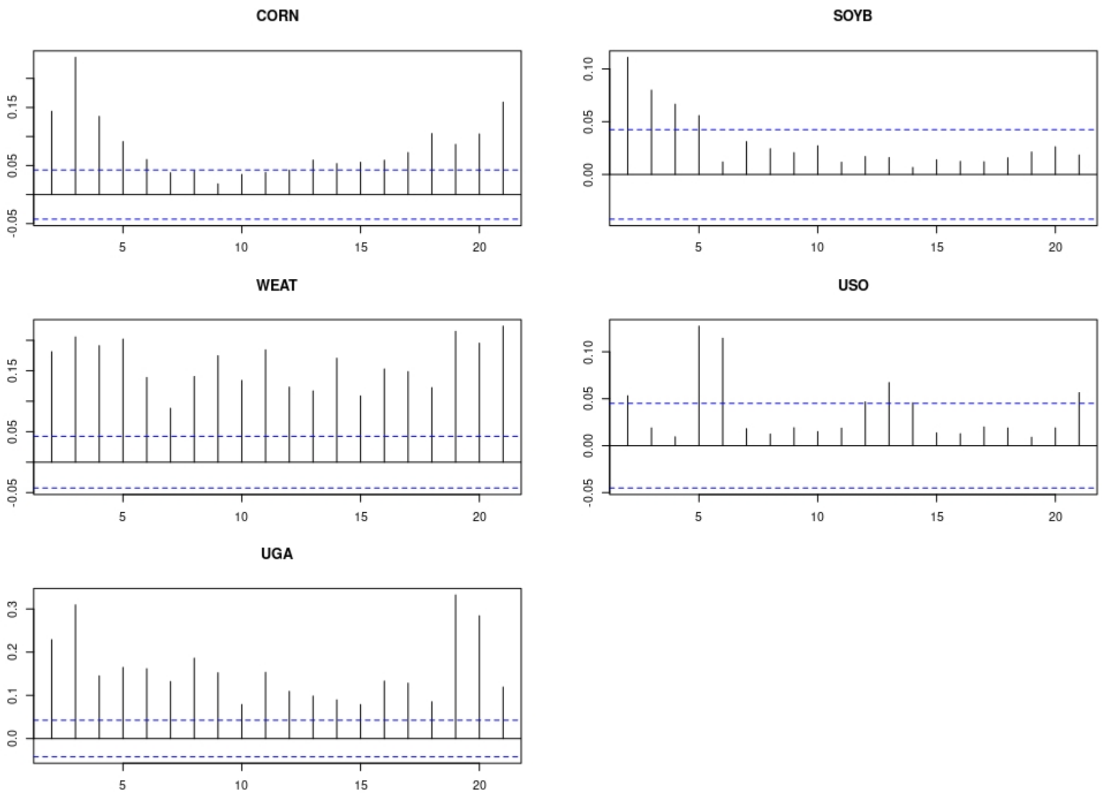

Figure 7: ACF Plot of Squared Non-systematic TD_A

Table 8: Ljung-Box Test Results

	x^2	P-value
CORN	652.6	< 0.01
SOYB	665.4	< 0.01
WEAT	1561.4	< 0.01
USO	584.5	< 0.01
UGA	1673.4	< 0.01

The presence of autocorrelation in the residuals of the Arbitrage Mincer-Zarnowitz regression implies a violation of the linear regression assumption of homoskedasticity in residuals (Wooldridge, 2013). To overcome this issue and better model variation in volatility, Engle (1982) proposed an approach wherein conditional variance is modeled as a linear function of previous residuals: the Autoregressive Conditional Heteroskedasticity (ARCH) model. The Generalized ARCH (GARCH) model developed by Bollerslev (1986) is a generalized extension of the ARCH model developed by Engle. The difference between these two models is analogous to the difference between an Autoregressive (AR) model and an Autoregressive Moving Average (ARMA) model wherein a GARCH model is able to more flexibly capture volatility clustering by including the previous estimation for conditional variance in the model (Bollerslev, 1986).

The Mincer-Zarnowitz Equation defined in Equation 8 can be extended to include a general form of the GARCH(p, q) model is as follows:

$$\begin{aligned} R_t^E &= \alpha + \beta R_t^N + \epsilon_t^{TD_A} \\ \epsilon_t^{TD_A} &\mid \psi_{t-1} \sim N(0, h_t) \\ h_t &= \omega + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^{TD_A 2} + \sum_{j=1}^{p} \delta_j h_{t-j} \end{aligned} \tag{12}$$

where p is greater than or equal to zero, q and ω are strictly positive, and γ_i and δ_i are greater than or equal to zero. The conditional variance h_t is thus a function of some constant ω, q lags of the squared residuals, and p lags of the previous estimates of conditional variance h. ψ_{t-1} is the information set at day $t-1$, in our case the previous residuals and the previous estimations for conditional variance.

While the original GARCH model assumes a normal distribution given an information set, other distributions can be used. Bollerslev (1987) proposed an extension of the GARCH model using a t-distribution to better capture changes in speculative asset prices. Where $p = q = 1$, the specification is as follows

$$\begin{aligned} R_t^E &= \alpha + \beta R_t^N + \epsilon_t^{TD_A} \\ \epsilon_t^{TD_A} &\mid \psi_{t-1} \sim T(0, h_t, \upsilon) \\ h_t &= \omega + \gamma \varepsilon_{t-1}^{TD_A 2} + \delta h_{t-1} \end{aligned} \tag{13}$$

where υ is the fitted degrees of freedom for the t-distribution and Γ is the Gamma Function. Examining the fitted distribution degrees of freedom (υ) yields information regarding the relative thickness of the distribution's tails. The smaller the υ estimate, the *fatter* the distribution's tails. The t-distribution approached the normal distribution as υ approaches infinity.

Ramos (2015) conducted a model comparison study of tracking differences in several developed and emerging market ETFs, finding that using a student-t distribution improves model fit compared to a normal distribution in the large majority of cases. We find that utilizing a t-distribution improves model fit for the ETFs studied.

GARCH models can also be expanded to include external variables which may explain changes in volatility. We include two variables. The first is the volatility of the ETF. The volatility of the ETF is perhaps the external variable which is most supported by previous empirical work, (see Aber et al., 2009; Shin and Soydemir, 2010; Lin and Chou, 2006; Fassas, 2015, for example). The economic motivation for including roll dates when modelling TE_A is that the arbitrage process becomes more difficult when markets move swiftly. Following the methodology common throughout literature (for example Shin and Soydemir (2010) and Aber et al. (2009)), we defined ETF volatility (σ) as the scaled intraday range of ETF prices as shown in Equation 14.

$$\sigma_t = \frac{High_t^{ETF} - Low_t^{ETF}}{Close_t^{ETF}} \tag{14}$$

The other variable included in the model is an indicator variable for roll dates, defined in Equation 15. Roll periods are a unique issue to futures-backed ETFs, commodity or otherwise. The motivation for including roll dates in modeling TE_A comes from analysis of the Creation and Redemption process. During the roll period, the Authorized Participant faces greater-than-normal uncertainty as to what the ETF manager will require or provide in exchange for ETF shares. Thus the Authorized Participant may be less willing to create or redeem shares during this period, increasing the volatility of TD_A by making is less attractive to keep the two prices aligned.

$$\text{Roll Date}_t = \begin{cases} 1 & \text{if } t = \text{Roll Date} \\ 0 & Otherwise \end{cases} \tag{15}$$

As discussed in the data section, rolling happens on a fixed schedule which varies across ETFs. While USO and UGA roll every month, CORN, SOYB, and WEAT roll just five times per year. The length of the roll period also varies across commodities with SOYB and UGA having single day roll periods, WEAT having 1-3 day roll periods, and CORN and USO having multiple day roll periods.

6 Model Results

Table 9 displays the model results for the base (no external regressors) and full (with external regressors) GARCH model specifications. Expanding the dataset to include roll dates (and other periods in the case of USO) does not significantly change the coefficient estimates for α and β. As in the original Mincer-Zarnowitz analysis, none of the α coefficients are statistically significantly. Additionally the β coefficients are all similiar, with CORN, USO, and UGA having β coefficients statistically different than one.

In all models, the γ and δ coefficients are significant, indicating that the GARCH model does explain some of the variation in volatility. Adding the two external regressors improves model fit based on the Akaike information criteria (AIC) for all ETFs with the exception of WEAT. In line with previous research, we find that the volatility of the ETF is a statistically significant contributor the volatility of residuals for CORN, SOYB, WEAT, and UGA. The positive coefficient values indicate that as the volatility of the ETF increases, so does TE_A.

We find no evidence that roll days (ξ) significantly effect TE_A in any of the ETFs studied. This lack of evidence is interesting given the uncertainty for the Authorized Participant around roll periods. To our knowledge, no other research has investigated this phenomenon.

Figure 8 plots the conditional variance estimates over time. SOYB and WEAT experienced significantly higher TE_A in 2012-2014. Two potential reasons for this decline is the volatility in the agricultural markets experienced at the beginning of the sample period, and the small AUM size of these ETFs during that time period. This size effect has been noted by Dorfleitner, Gerl, and Gerer (2018) and Chu (2011). The large spike seen in USO and UGA is associated with volatility in the energy markets due to the COVID-19 pandemic. On April 20, 2020, the May 2020 WTI crude oil contract reached a low of -\$37.63. Though USO did not hold this contract at the time, there was significant issues with the ETF, prompting the fund managers to change the ETF benchmark and eventually complete a 8-1 reverse -split on April 28, 2020 (USCF, 2020). This may have contributed to additional TE_A above what would solely been expected due to volatility. It is also interesting to note that the agricultural ETFs seem to have also experienced higher-than-average TE_A at this time, despite not necessarily experiencing increased market volatility due to the pandemic.

The fitted values for υ give some indication of the *fattness* of the distributions tails. Based on their smaller υ values, SOYB, USO, and UGA have fatter tails that CORN and WEAT. All ETFs have significantly fatter tails than a normal distribution. This is further evidence that the Student t-distribution is preferable when modelling Tracking Error compared to a Normal distribution.

Overall, the results of our model are largely in line with previous research with respect to volatility being a contributor to TE. Our results further highlight the need to incorporate the time-varying nature and non-normality when analyzing Tracking Errors. These attributes have important implications when modeling.

Table 9: GARCH Model Results

	CORN		SOYB		WEAT		USO		UGA	
	Base	Full	Base	Full	Base	Full	Base	Full	Base	Full
α	$-5.88E-04$ (0.0038)	$-9.66E-04$ (0.0036)	$-5.57E-05$ (0.0042)	$-2.50E-04$ (0.0042)	$-3.34E-04$ (0.0065)	$-3.10E-04$ (0.0065)	$3.56E-04$ (0.0068)	$-4.33E-04$ (0.0065)	$5.28E-05$ (0.0060)	$-7.48E-04$ (0.0059)
β	**0.9795**** (0.0039)	**0.9784**** (0.0041)	0.9998 (0.0049)	0.9994 (0.0051)	0.9925 (0.0050)	0.9926 (0.0050)	**0.9414**** (0.0054)	**0.9454**** (0.0053)	**0.9488**** (0.0044)	**0.9462**** (0.0047)
ω	**0.0115**** (0.0034)	$4.47E-04$ (0.0022)	**0.0142**** (0.0026)	0.0038 (0.0028)	**0.0035**** (0.0012)	0.0012 (0.0022)	**0.0167**** (0.0045)	$1.37E-13$ ($9.70E{-}05$)	**0.0286**** (0.0055)	$1.84E-10$ ($1.13E{-}04$)
δ	**0.2821**** (0.0501)	**0.2960**** (0.0397)	**0.4534**** (0.0620)	**0.4461**** (0.0592)	**0.1164**** (0.0264)	**0.1239**** (0.0290)	**0.3832**** (0.0614)	**0.4482**** (0.0574)	**0.5075**** (0.0642)	**0.4619**** (0.0500)
γ	**0.5221**** (0.0953)	**0.5480**** (0.0744)	**0.5035**** (0.0501)	**0.4514**** (0.0523)	**0.8663**** (0.0289)	**0.8551**** (0.0330)	**0.6158**** (0.0544)	**0.3157**** (0.0574)	**0.4470**** (0.0566)	**0.2006**** (0.0520)
σ		**0.0169**** (0.0024)		**0.0129**** (0.0031)		0.0017 (0.0014)		**0.0327**** (0.0047)		**0.0396**** (0.0047)
ξ		$6.23E-10$ (0.0127)		$1.36E-11$ (0.0138)		$4.37E-08$ (0.0055)		$4.85E-10$ (0.0128)		$5.12E-09$ (0.0215)
υ	10.49	18.7654	4.58	4.8332	8.68	8.7612	5.6492	6.4486	7.22	10.4717
AIC	-0.1913	-0.2336	0.3315	0.3204	0.9159	0.9170	1.2227	1.1666	1.0612	1.0015

Model results from GARCH modeling of Tracking Errors. Mean Equation is $R^E = \alpha + \beta R^N + \epsilon$. Variance equation differs. Base Model is solely GARCH (1,1) while Full model includes external variables: ETF Volatility and Roll Dates.

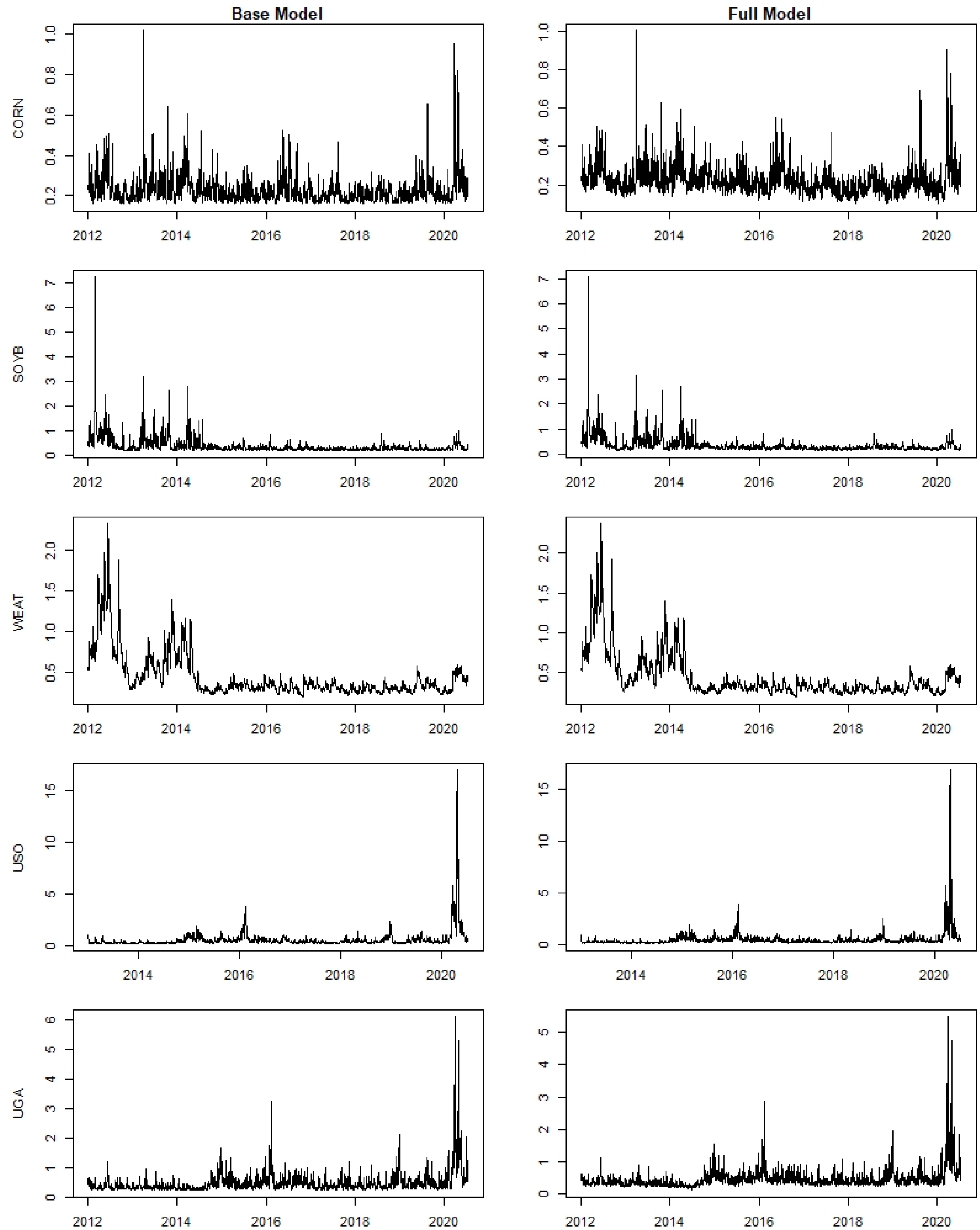

Figure 8: Estimated Conditional Variance for Base and Full GARCH Models

7 Conclusions

We investigated the tracking divergence in five of the most popular futures-backed commodity ETFs. By decomposing total tracking differences, we are able to properly identify the relative contributions to tracking differences of the ETF managers and the arbitrage process. Our research offers three contributions to the research base.

The first key contribution is our findings on the relative size of Managerial and Arbitrage Tracking Difference. Because of the difficulty in reconstructing asset baskets and the relative nichness of futures-backed commodity ETFs, no papers, to our knowledge, investigate Managerial Tracking Differences for these ETFs with the exception of Neff and Isengildina-Massa (2018). Contrary to previous literature investigating equity ETFs, we find that for all of the ETFs studied, managers do a relatively good job of keeping the NAV inline with the benchmark asset basket compared to the Creation and Redemption process. For the Agricultural ETFs (CORN, SOYB, and WEAT), very rarely does daily managerial tracking difference exceed 10 basis points. For the Energy ETFs (USO and UGA), the performance is even better, with daily managerial tracking differences generally less than 3 basis points. The differences between ETF Price and NAV are significantly greater for all of the ETFs studied.

The second contribution is our analysis of average tracking differences, applying the Mincer-Zarnowitz technique to the decomposition of total tracking differences to gain insights. We find that for CORN, SOYB, and WEAT, there is a negative average difference between the return of the NAV and the benchmark, likely due in part to the large stated expense ratio of the fund. For CORN, USO, and UGA, we find that the variation in NAV is not transferred completely to the ETF price: the volatility of the ETF is lower than the volatility of the NAV. This is especially true for USO and UGA.

Our third contribution to the research body is the findings related to the dynamics of Tracking Error: the volatility of tracking differences. We focus on the Tracking Error attributable to the Arbitrage process. As in previous studies, we find support that tracking error is non-constant. We test two external regressors: the volatility of the ETF and roll periods. Rolling periods are unique to futures-backed ETFs and are understudied. We find no evidence of roll periods contributing to changes in TE_A. As in previous literature, we find evidence that the volatility of the ETF effects TE_A.

Overall, the findings of our research highlight the heterogeneity of tracking success amongst ETFs, even a small subsection such as futures-backed commodity ETFs. ETF investors should be aware of the issues that the ETFs have in achieving their primarily goal of replicating benchmark exposure, especially in times of market volatility. Futhermore, our results indicate that any efforts to improve the tracking ability of the five ETFs studied should focus on the Creation and Redemption process, as that is the primary source of tracking divergence in all of the ETFs studied.

References

Aber, J.W., D. Li, , and L. Can. 2009. "Price volatility and tracking ability of ETFs." *Journal of Asset Management* 10:210–221.

Bloomberg. 2016. "The Bloomberg Commodity Index: A Primer on Index Calculation and Performance." https://data.bloomberglp.com/indices/sites/2/2016/01/BCOM-Calculation-Primer.pdf.

Bollerslev, T. 1987. "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *The Review of Economics and Statistics* 69:542–547.

—. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–327.

Chu, P.K.K. 2011. "Study on the tracking errors and their determinants: evidence from Hong Kong exchange traded funds." *Applied Financial Economics* 21:309–315.

Dickey, D.A., and W.A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74:427–431.

Dorfleitner, G., A. Gerl, and J. Gerer. 2018. "The pricing efficiency of exchange-traded commodities." *Review of Managerial Science* 12:255–284.

Elton, E.J., M.J. Gruber, G. Comer, and K. Li. 2002. "Spiders: Where Are the Bugs?" *The Journal of Business* 75:453–472.

Engle, R., and D. Sarkar. 2006. "Premiums-Discounts and Exchange Traded Funds." *The Journal of Derivatives* 13:27–45.

Engle, R.F. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987–1008.

ESMA. 2012. *Guidelines for competent authorities and UCITS management companies*. : European Securities and Markets Authority, ESMA/2012/832EN, 12.

Fassas, A.P. 2015. "Tracking Ability of ETFs: Physical versus Synthetic Replication." *The Journal of Index Investing* 5:9–20.

Frino, A., and D.R. Gallagher. 2001. "Tracking S&P 500 Index Funds." *Journal of Portfolio Management* 28.

Gallagher, D.R., and R. Segara. 2006. "The performance and trading characteristics of exchange-traded funds." *Journal of Investment Strategy* 1:49–60.

Gastineau, G.L. 2010. *The Exchange-Traded Funds Manual*, 2nd ed. John Wiley Sons, Inc.

Gorton, G., and K.G. Rouwenhorst. 2004. "Facts and Fantasies about Commodity Futures." Working Paper No. 10595, National Bureau of Economic Research, June.

Guedj, I., G. Li, and C. McCann. 2011. "Futures-Based Commodities ETFs.", pp. .

Hill, J., D. Nadig, and M. Hougan. 2015. *A Comprehensive Guide to Exchange-Traded Funds (ETFs).*

IPE. 2019. "2019 Exchange-Traded Funds Guide." Industry paper, Investments and Pensions Europe.

Irwin, S.H., and D.R. Sanders. 2012. "Financialization and Structural Change in Commodity Futures Markets." *Journal of Agricultural and Applied Economics* 44:371–396.

Irwin, S.H., D.R. Sanders, A. Smith, and S. Main. 2020. "Returns to Investing in Commodity Futures: Separating the Wheat from the Chaff." *Applied Economic Perspectives and Policy* 00:1–28.

Johnson, B., H. Bioy, A. Kellett, and L. Davidson. 2013. "On The Right Track: Measuring Tracking Efficiency in ETFs." White paper, Morning Star.

Lin, A., and A. Chou. 2006. "The Tracking Error and Premium/Discount of Taiwan's First Exchange Traded Fund." *Web Journal of Chinese Management Review* 9.

Ljung, G., and G. Box. 1978. "On a Measure of a Lack of Fit in Time Series Models." *Biometrika* 65:297–303.

Mincer, J.A., and V. Zarnowitz. 1969. *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, National Bureau of Economic Research, chap. The Evaluation of Economic Forecasts. pp. 3–46.

Neff, T., and O. Isengildina-Massa. 2018. "How Well do Commodity ETFs Track Underlying Assets?" Working paper, Virginia Tech.

Perera, D., J. Bialkowski, and M.T. Bohl. 2018. "Is the tracking error time varying? Evidence from agricultural ETCs." Working paper, University of Canterbury, Canterbury, New Zealand.

Qadan, M., and J. Yagil. 2012. "On the dynamics of tracking indices by exchange traded funds in the presence of high volatility." *Managerial Finance* 38:804–832.

Ramos, J.a.P.M. 2015. "Tracking Ability of Global Emerging Markets Exchange Traded Funds." MS thesis, ISCTE Business School.

Shin, S., and G. Soydemir. 2010. "Exchange-traded funds, persistence in tracking errors and information dissemination." *Journal of Multinational Financial Management* 20:214–234.

Teucrium. 2020. "Teucrium Fund Prospectus." October 2, https://teucrium.com/etfs/corn.

USCF. 2020. "UNITED STATES OIL FUND LP Prospectus." June 12, https://www.uscfinvestments.com/documents/united-states-oil-fund-pro-20200612.pdf.

Wooldridge, J.M. 2013. *Introductory Econometrics: A Modern Approach*, 5th ed. South-Western Cengage.

Appendices

A Appendix

A.1 Creation Redemption Process

Simplified Creation and Redemption Process: Teucrium Funds (CORN, SOYB, WEAT) (Teucrium, 2020)

1. Irrevocable creation order placed by Authorized Participant (AP) before 1:15pm EST
2. End of Day (4:00pm EST): The Fund Sponsor sets the creation/redemption basket, determining the cash, cash equivalents, and/or commodity futures, including the maturities of those cash equivalents, which can be exchanged for shares.
3. Purchase Settlement Date (Normally end of the following day): The AP transfers the Custodian the ETF shares (creation basket) and receives the redemption basket (ETF shares).

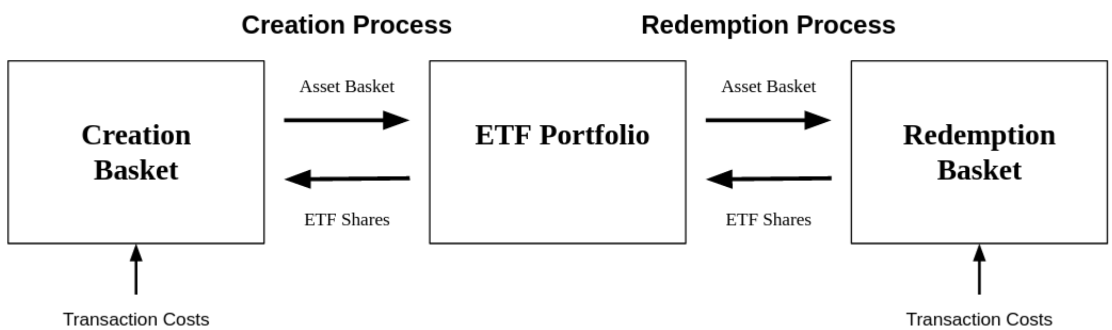

Figure 9: Creation Redemption Process. Adapted from Gastineau (2010)

A.2 Expanded Dataset Summary Statistics

Table 10: Updated ETF Price and NAV Summary Statistics inclusive of Roll Dates

	Value	CORN	SOYB	WEAT	USO	UGA
ETF	Min	11.67	13.34	4.86	2.13	8.90
	Median	21.27	19.00	8.88	14.31	32.57
	Mean	24.93	19.68	10.99	20.56	38.62
	Max	52.67	28.85	25.35	42.01	65.71
	Std Dev	10.26	3.47	5.76	11.18	14.12
NAV	Min	11.70	13.38	4.88	2.04	8.52
	Median	21.26	19.00	8.88	14.30	32.60
	Mean	24.93	19.68	10.98	20.56	38.62
	Max	52.68	28.77	25.29	42.00	65.48
	Std Dev	10.26	3.48	5.75	11.19	14.13
	N	2,143	2,137	2,137	2,143	2,143

Table 11: Updated ETF and NAV Return Statistics inclusive of Roll Dates

	Value	CORN	SOYB	WEAT	USO	UGA
ETF	Min	-7.26	-7.55	-6.53	-29.19	-25.29
	Median	-0.05	0.00	-0.14	0.03	0.00
	Mean	-0.06	-0.02	-0.07	-0.01	-0.04
	Max	6.34	9.22	7.72	213.42	17.95
	Std Dev	1.23	1.15	1.51	5.20	2.25
NAV	Min	-6.19	-5.77	-5.96	-51.90	-26.52
	Median	-0.05	-0.01	-0.12	0.03	0.03
	Mean	-0.06	-0.02	-0.07	-0.01	-0.04
	Max	6.59	4.54	6.82	213.12	19.30
	Std Dev	1.23	1.03	1.44	5.37	2.43
	N	2,142	2,136	2,136	2,142	2,142

Table 12: TD_A Expanded Dataset Summary Statistics

	Value	CORN	SOYB	WEAT	USO	UGA
TDa	Min	-1.22	-7.90	-4.21	-22.97	-7.78
	Median	-4.10E-04	-6.00E-03	2.64E-03	-6.24E-03	8.53E-04
	Mean	3.02E-05	1.18E-04	4.10E-04	-1.16E-04	-1.75E-04
	Max	1.75	9.00	4.23	23.05	8.20
	STD DEV	0.24	0.51	0.53	1.06	0.64
	N	2142	2136	2136	1891	2142