THE MEASUREMENT OF TRACKING ERRORS OF GOLD ETFS: EVIDENCE FROM CHINA

Wei-Fong Pan[1*], Ting Li[2]

1. Investment Analyst, Sales and Trading Department, Ping An Futures Co. Ltd. China
2. Portfolio Manager, Sino Life Asset Management Co. China

* Corresponding Author: Wei-Fong Pan, Unit 01, 14/F Rong Chao Building, No. 4036, Jintian Road, Futian District, Shenzhen, China.
☎ +86(0)75522628910   ✉ panweifeng550@pingan.com.cn

Abstract: This paper presents the first study on the measurement of tracking errors using daily figures for gold exchange-traded funds (ETFs) in China. Three methods are employed to measure tracking errors: 1) calculating the absolute error measure, 2) calculating the differences between the standard deviation of the benchmark index and the ETF, and 3) a regression analysis of empirical returns. In general, the results suggest that the tracking errors of these ETFs in China are lower than those of equity-based ETFs in Hong Kong, the US, and Australia. This study further applied two optimised replication portfolios (50-10-10-30 and 90-2-3-5) for a total of three types of simulation portfolio. The overall results suggest that the performances of the optimised replication portfolios were better than the performance of the full replication portfolio. Our results provide valuable insight for both institutional and retail investors and the opportunity for exposure to a wide range of commodity ETFs in China.

Keywords: Tracking Errors; ETFs; Commodities; Gold; China

1.   Introduction

The development of exchange-traded funds (ETFs) provides opportunities for both institutional and retail investors to be exposed to a wide range of asset classes. A bulk of the existing studies focuses on the tracking errors of equity-based ETFs using distinct approaches. Using S&P 500 Index data, Frino et al. (2004) examined the exogenous determinants of tracking errors and observed that such errors are significantly influenced by index revisions, share issuances, spin-offs, share repurchases, index replication strategy, and fund size. They also found a seasonal pattern in tracking errors, consistent with the finding of Frino and Gallagher (2001). Chu (2011) studied the magnitude and determinants of ETF tracking errors using daily data on the Hong Kong stock market and found that the tracking error in Hong Kong is higher than those in the United States and Australia. Avellaneda and Zhang (2010) studied the price behaviour equity-leveraged ETFs in different sectors and found minimal one-day tracking errors among the most liquid equity ETFs.

Commodities are unique in part because physical assets cannot be stored easily owing to the extra costs for warehousing. Thus, futures-based commodity ETFs may fail to track their reference indices perfectly. The commodity is also counter-cyclical with stocks and bonds; studies observed that it is significantly negatively correlated with both bonds and equities, implying that an appropriate allocation to commodities enhances portfolio performance (Jensen et al., 2002; Fuertes et al., 2010). Gold is often viewed by investors as a hedge against market turmoil. A typical example of this tendency is when the price of gold was pushed to an all-time high of US$1900 in August 2011 owing to the global financial crisis and European sovereign debt crisis at the time.[1] Although commodities, especially gold, are important both for risk hedging and for asset management, there are few studies of tracking errors in commodity ETFs.[2]

The contribution of this study is to measure the tracking errors of commodity ETFs and use optimised way to replicate the return of commodity ETFs in China, from 05 Jan 2015 to 29 Feb 2016. To the best our knowledge, this study is the first to investigate all four existing gold ETFs in China. Existing studies paid more attention to equity-based ETFs in either the United States or European countries rather than in emerging countries. Following Pope and Yadav (1994) and Shin and Soydemir (2010), this study employed three different approaches to estimate tracking errors in order to obtain robust results.

The rest of the paper is organised as follows. Section 2 presents the data sources and provides an overview of the development of commodity ETFs in China. Section 3 describes the empirical approaches used to estimate the tracking error. Section 4 discusses the empirical findings, and Section 5 provides the conclusions and some directions for future research.

2.   The Development of Commodity ETFs in China

The development of gold ETFs enables investors to allocate some of their assets to gold without directly buying physical gold. Gold ETFs in China first emerged on 24 June 2013, developed by GuoTai Fund Management Company, and the country has since become the largest gold consumer in the world. The Shanghai Gold Exchange facilitates spot gold exchange.

Table 1 shows that the trading volume of spot gold in China has significantly increased along with the trading amount, suggesting that investors have become more focused on gold investments, which, in turn, makes this study important and timely.

The commodity ETFs used in this study are HuaAn Gold ETF, GuoTai Gold ETF, Bosera Gold ETF, and E Fund Gold ETF.[3] The ETF prices were collected from the Wind Database, created by Wind Information Co., Ltd., a financial data provider in China. Since the commodity ETFs in China emerged later than those in developed countries, all four commodity ETFs track the gold spot price at the Shanghai Gold Exchange, which is also

---

[1] Białkowski et al. [2015] investigated the gold price during these crises.
[2] Some of these few existing studies are those by Murphy and Wright [2010], Guedj et al. [2011], and Leung and Ward [2015].
[3] Our dataset excludes the UBS SDIC Silver LOF (listed-open fund), first, because it has been traded for less than half a year, and second, LOFs differ from ETFs in some aspects, such as in the redemption mechanism.

**Table 1: Trends in Gold Trading in Shanghai Gold Exchange, 2006-2014**

	Trading Volume (ton)	Trading Amount (0.1 billion)
2006	1,249.60	1,947.51
2007	1,828.13	3,164.90
2008	4,463.77	8,696.05
2009	4,710.82	10,288.76
2010	6,051.50	16,157.81
2011	7,438.45	24,772.00
2012	6,350.20	21,506.00
2013	11,614.00	32,134.00
2014	18,486.00	45,900.00

Source: Shanghai Gold Exchange

the source of the gold spot price in this study. All of the data reflect daily observations for each trading day from 05 Jan 2015 to 29 Feb 2016.

Figure 1 shows the performance of the existing gold ETFs in China. All four ETFs show a similar trend, with very small variations, and have a net asset value (NAV) between 2.00 and 2.65. However, even such small variations would have a large impact on the ETF returns.

Figure 1: Performance of the four gold ETFs in China, Mar 2015-2016
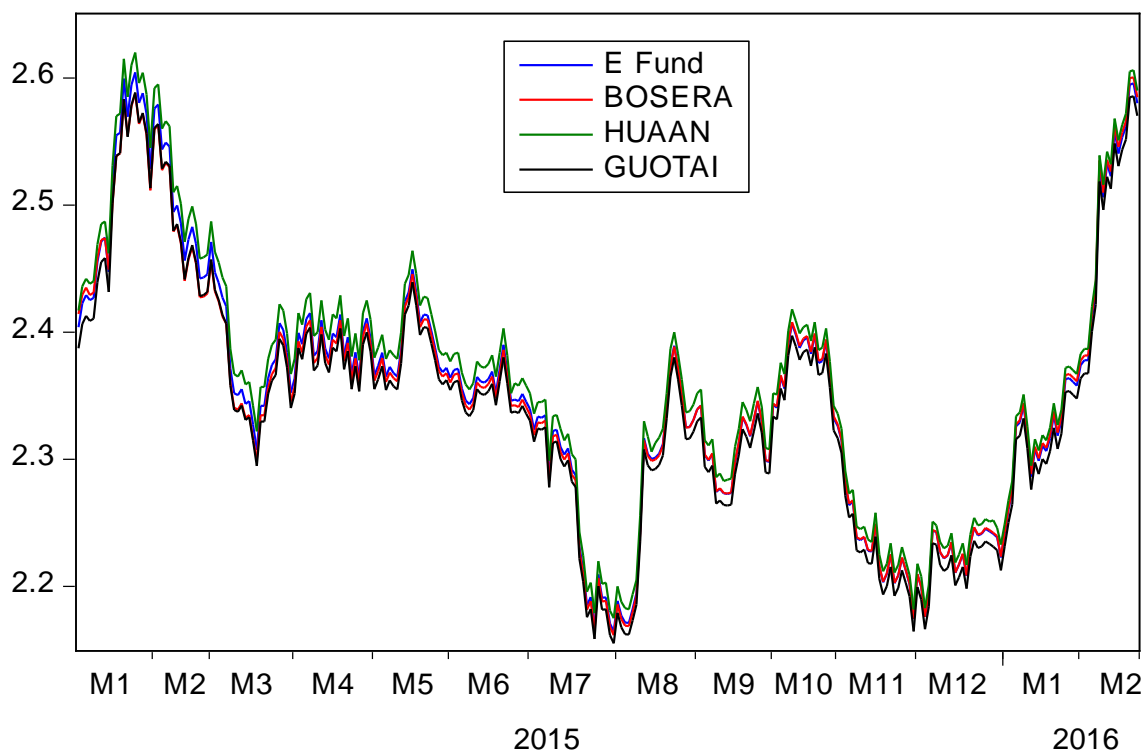
3.   Empirical Methodology

This section reviews the possible sources of tracking errors and the methods for analysing such errors. The tracking error, ceteris paribus, is zero if the index fund perfectly aligns with the benchmark index. However, in practice, an ETF's performance in tracking the index is affected by a few factors, such as management fees and administrative/ operating expenses, different compositions of the index fund and the index, and trading costs. (Frino and Gallagher, 2001; Drenovak et al., 2014). Thus, the tracking error is non-zero in practice, as was observed by many empirical studies (see for example, Murphy and Wright, 2010).

Several articles explored important issues in tracking error measurement. Roll (1992) provided a criterion for analysing ETF performance. The approaches for tracking error estimation were well documented in the academic literature (e.g. Pope and Yadav, 1994; Shin and Soydemir, 2010). This study employs three methods to measure the tracking errors. One of the traditional methods involves calculating the absolute error measure, which is defined as the average absolute value of the difference between the returns of the benchmark index and index fund. The measure can be described as follows:

$$\text{TE}_{1,i} = \frac{\left(\sum_{i=1}^{n}|R_{f,t}-R_{x,t}|\right)}{n}, \quad t = 1,2,3 \dots n \tag{1}$$

where $R_{f,t}$ represents the return of index fund f at time t, while $R_{x,t}$ is the return of its underlying gold at time t.

The second method of tracking error estimation involves calculating the standard deviation of the difference in returns of the benchmark index and that of the ETF. The variance equation can be described as follows:

$$\text{TE}_{2,i} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[(R_{f,t}-R_{x,t})-\frac{1}{n}\sum_{i=1}^{n}(R_{f,t}-R_{x,t})\right]^2} \tag{2}$$

where t denotes the time period. $R_{f,t}$ represents the return of index fund f at time t, while $R_{x,t}$ is the return of its underlying (Gold) at time t. We can rewrite equation (2) as:

$$\text{TE}_{2,i} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[e_{f,t}-\bar{e}_x\right]^2} \tag{3}$$

where $e_{f,t} = R_{f,t} - R_{x,t}$

The third method of tracking error estimation involves a regression analysis of empirical returns, based on the following linear model:

$$R_{f,t} = \alpha + \beta R_{x,t} + \varepsilon \tag{4}$$

where $\varepsilon \sim N(0, \sigma^2)$ is the error term. The tracking error is defined as the standard error of equation (4). In the case of ETFs pursuing a passive investment strategy, α is not expected to be statistically different from zero, while β is not expected to be statistically different from one. A very high R^2 is also expected.

4. Results

We begin with estimating the tracking error using the absolute error method (TE1). From Figure 2, which presents the TE1 variation of all the gold ETFs, it is clear that the highest tracking error occurs on Jan 2015. For robustness, we used three samples according to time period—the full sample (Jan 2015-Mar 2016), a sample of only one year during the study period (Mar 2015-2016), and a sample only of the last six months of the study period (Sept 2015-Mar 2016). Table 2 reports the empirical results of the tracking error estimation using the three methods.

We first consider the full sample, that is, the sample for the entire study period. The daily tracking error based on the first estimation method (calculating the absolute error measure) (TE1) ranges from 0.0024% to 0.0273% across all ETFs. The daily tracking error based on the second method (calculating the standard deviation of return differences) (TE2) ranges from 0.0035 % to 0.05%. Meanwhile, the daily tracking error based on the third method (regression analysis of empirical returns) (TE3) ranges from 0.0027% to 0.0499%, and the coefficient of the benchmark index, as expected, is very close to one and the R^2 and is nearly 100%. The tracking error of the gold ETFs in China is generally lower than those of equity-based index ETFs in Hong Kong (0.39%), Australia (0.0074%), and the United States (0.039%) (Chu, 2011). The measures from all three methods indicate that HuaAn Gold ETFs have the highest tracking error among four ETFs and GuoTai Gold ETFs have the best performance and the smallest tracking error.

We first consider the full sample, that is, the sample for the entire study period. The daily tracking error based on the first estimation method (calculating the absolute error measure) (TE1) ranges from 0.0024% to 0.0273% across all ETFs. The daily tracking error based on the second method (calculating the standard deviation of return differences) (TE2) ranges from 0.0035 % to 0.05%. Meanwhile, the daily tracking error based on the third method (regression analysis of empirical returns) (TE3) ranges from 0.0027% to 0.0499%, and the coefficient of the benchmark index, as expected, is very close to one and the R^2 and is nearly 100%. The tracking error of the gold ETFs in China is generally lower than those of equity-based index ETFs in Hong Kong (0.39%), Australia (0.0074%), and the United States (0.039%) (Chu, 2011). The measures from all three methods indicate that HuaAn Gold ETFs have the highest tracking error among four ETFs and GuoTai Gold ETFs have the best performance and the smallest tracking error.
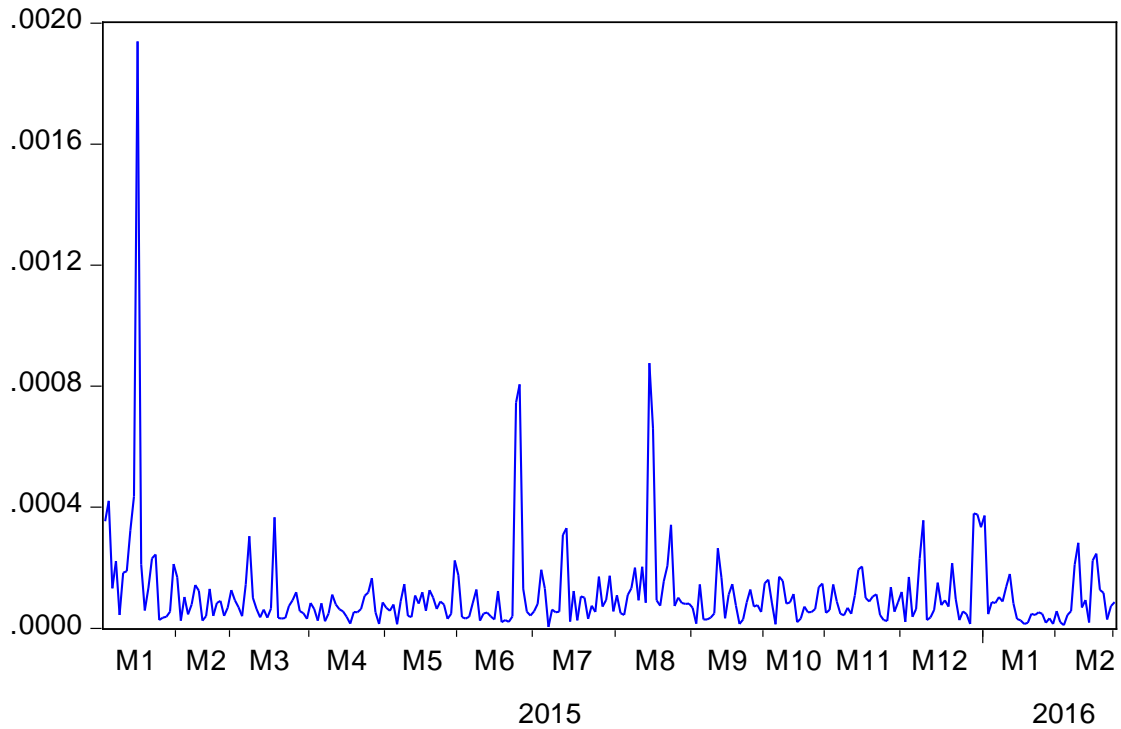
Figure 2: Average Gold ETF Tracking Errors (TE1)

Table 2: Tracking Errors of Gold ETFs for the Three Samples

Full Sample Period										
	Absolute Error				Return Differences		Regression Analysis			
	TE_1	(%)			TE_2 (%)		TE_3			
	Mean	SD	Min	Max	SD	Mean	ε	α	β	R^2(%)
GuoTai	0.0024	0.0026	0.00	0.0249	0.0035	0.0002	0.0035	0.0002	0.9996	99.999
HuaAn	0.0273	0.0417	0.00	0.3449	0.0500	-0.0013	0.0499	-0.0012	0.9983	99.692
Bosera	0.0128	0.0463	0.00	0.7089	0.0481	-0.0015	0.0466	-0.0010	0.9866	99.724
E Fund	0.0033	0.0029	0.00	0.0151	0.0100	-0.0009	0.0027	-0.0008	0.9962	99.999
Mar 2015-2016										
GuoTai	0.0023	0.0025	0.00	0.0249	0.0034	-0.0001	0.0034	-0.0001	0.9998	99.999
HuaAn	0.0286	0.0439	0.00	0.3449	0.0523	-0.0018	0.0525	-0.0018	1.0000	99.644
Bosera	0.0072	0.0140	0.00	0.1372	0.0156	0.0024	0.0155	0.0025	0.9979	99.969
E Fund	0.0033	0.0029	0.00	0.0151	0.0043	-0.0009	0.0028	-0.0008	0.9963	99.999
Sep 2015-Mar 2016										
GuoTai	0.0021	0.0026	0.00	0.0249	0.0033	-0.0004	0.0033	-0.0003	0.9998	99.999
HuaAn	0.0272	0.0312	0.00	0.1453	0.0414	-0.0014	0.0415	-0.0011	0.9961	99.782
Bosera	0.0066	0.0115	0.00	0.083	0.0132	0.0013	0.0132	0.0011	1.0017	99.978
E Fund	0.0029	0.0026	0.00	0.0125	0.0039	-0.0005	0.0027	-0.0002	0.9969	99.999

Note: Tracking errors are expressed as percentages.

For robustness, we also consider two other samples—one of only one year of the study period (Mar 2015-2016) and another of only the last six months of the study period (Sept 2015-Mar 2016). TE1 is between 0.0032 and 0.0286 across all ETFs, TE2 is between 0.0034 and 0.0523, and TE3 is between 0.0028 and 0.0525. For the full sample period, the order of the ETFs in terms of the magnitude of the tracking error is the same when using the first two methods. However, the results of the third method (regression analysis of empirical returns) show that E Fund Gold ETFs, not GuoTai Gold ETFs, have the lowest tracking error. A similar situation is observed for the sample for the last six months of the study period—the results of the first two methods suggest that GuoTai Gold ETFs have the smallest tracking error, but the results of the third method show that E Fund Gold ETFs have the best performance. The study's results support Pope and Yadav's (1994) idea that if β is not exactly equal to one, the order of the ETFs in terms of the magnitude of the tracking error may be different. Pope and Yadav (1994) also pointed out that if the relationship between the benchmark index return and Gold ETF return is not linear, the third method may overestimate the tracking error.

Finally, currently in Shanghai, there are five firm contracts for spot gold trading: AU99.99, AU99.95, AU99.5, AU100g, and AU50g; this study uses AU99.95 as its base to construct one full replication portfolio and two optimised replication portfolios for a total of three types of simulation portfolios in order to test the empirical results. For the empirical comparison, the study uses one 95-5 full replication portfolio (95% spot gold and 5% cash), and two types of simulation portfolios: one optimised replication portfolio with a 50-10-10-30 composition (50% spot gold, 10% gold deferred settlement contracts, 10% cash, and 30% monetary funds), and another optimised replication portfolio with a 90-2-3-5 composition (90% spot gold, 2% gold deferred settlement contracts, 3% cash, and 5% monetary funds). These portfolios are structured in this way primarily because of current mainstream portfolio construction principles. The 95-5 composition of the full replication portfolio follows basic current market principles for ETFs, whereby the portfolio contains 95% of the physical product, plus 5% cash to respond to redemption application demand. For the 50-10-10-30 construction, the 10% ratio of gold deferred settlement contracts considers current investment fund companies' highest proportion of derivative default investments under an upper-limit scenario. The 90-2-3-5 construction considers the weighted results of the lower limit of an index fund's physical position and the security amplification effect. Because gold deferred settlement contracts may collect deferred payments, and their cost is associated with the number of contracts settled on a given day, when settlements are equal in number to those declared, no delayed compensation payments are made. When settlements are not equal in number to those declared, the side with fewer declarations makes delayed compensation payments to the side with greater declarations. The detailed data of daily settlement declarations are not obtainable and are relatively random; thus, when constructing portfolios, the effects of delayed compensation payments are not considered.

From the three scenarios described above, considering the full sample tracking errors (Table 3), the 90-2-3-5 portfolio had significantly lower tracking errors than the two other portfolios. Considering annual tracking errors, the 90-2-3-5 portfolio had the smallest tracking errors and was the most stable. The 50-10-10-30 portfolio had relatively large annual tracking errors due to fluctuations resulting from the security amplification effect. From this perspective, the 90-2-3-5 portfolio was slightly better.

Table 3: Average Tracking Error of Replications (%)

Time Period	Full replication	90-2-3-5	50-10-10-30
Overall	0.0155	0.0115	0.0186
Mar 2015-Mar 2016	0.0189	0.0161	0.0257
Sep 2015-Mar 2016	0.0104	0.0097	0.0152

5. Conclusion

ETFs have provided both institutional and retail investors with new opportunities to be exposed to a wide array of commodities. This study is the first to examine the measurement and determinants of tracking errors using daily data for gold ETFs in China from Jan 2015 to Mar 2016. The study's results show that the tracking error of gold ETFs is generally lower than those of equity-based ETFs in Hong Kong, the United States, and Australia. The results consistently indicate HuaAn Gold ETFs have the highest tracking error among all the four gold ETFs. The results also support Pope and Yadav's (1994) finding that the tracking error calculated from a regression analysis may differ from the standard deviation of return difference if the coefficient of the benchmark index is not exactly equal to one.

This study further applied an analysis of five types of gold target products to establish a full replication portfolio and two optimised replication portfolios (50-10-10-30 and 90-2-3-5) for a total of three types of simulation portfolio in order to test the empirical results. The overall results suggest that the performances of the optimised replication portfolios were better than the performance of the full replication portfolio. Specifically, the 90-2-3-5 optimised replication portfolio having both the lowest tracking error and better performance than the 50-10-10-30 optimised replication portfolio in the event of shocks or a fall in the market.

These findings provide important information for investors, particularly in terms of the measurement of commodity ETFs in China. However, some limitations still remain. When testing the simulated portfolios, it was assumed: 1) there were no redemption applications for ETFs, 2) purchases were redeemed within each day for the ETFs, and, 3) an 'odd lot' issue existed when real purchase redemption assets could not be incorporated in whole integer multiples into the optimised replication simulated portfolio's gold deferred payment contracts and AU99.95 portfolios. This could lead to increases in tracking errors; thus, how to resolve the odd lot issue should be addressed in future research. This study's framework could be extended to investigate other types of ETFs, particularly, in the context of other countries, and to examine other determinants of tracking errors in commodity ETFs.

References

Avellaneda M. and Zhang S. (2010). Path-dependence of Leveraged ETF Returns. SIAM Journal of Financial Mathematics, 1, 586–603.

Białkowski, J., Bohl, M. T., Stephan, P. M. and Wisniewski, T. P. (2015). The Gold Price in Times of Crisis. International Review of Financial Analysis, 41, 329-339.

Frino, A. and Gallagher, D. R. (2001). Tracking S&P 500 Index Funds. Journal of Portfolio Management, 28, 44-55.

Frino A., Gallagher D.R., Neubert A.S. and Oetomo T.N. (2004). Index Design and Implications for Index Tracking: Evidence from S&P 500 Index Funds. Journal of Portfolio Management, 88-95.

Fuertes, A., Miffre, J. and Rallis, G. (2010). Tactical Allocation in Commodity Futures Markets: Combining Momentum and Term Structure Signals. Journal of Banking & Finance, 34, 2530- 2548.

Grinblatt, M. and Titman, S. (1989). Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings. Journal of Business, 62, 393–416.

Guedj I., Li G. and McCann C. (2011). Futures-based Commodities ETFs. The Journal of Index Investing, 2, 14–24.

Jensen, G. R., Johnson, R. R., and Mercer, J. M. (2002). Tactical Asset Allocation and Commodity Futures. The Journal of Portfolio Management, 28, 100-111.

Leung T. and Ward B. (2015). The Golden Target: Analyzing the Tracking Performance of Leveraged Gold ETFs. Studies in Economics and Finance, 32, 278-97.

Murphy R. and Wright C. (2010). An Empirical Investigation of the Performance of Commodity-based Leveraged ETFs. The Journal of Index Investing, 1, 14–23.

Chu. P.K. (2011). Study on the Tracking Errors and Their Determinants: Evidence from Hong Kong Exchange Traded Funds. Applied Financial Economics, 2, 309-315.

Pope, P. and Yadav, P. (1994). Discovering Errors in Tracking Error. Journal of Portfolio Management, 20, 27–32.

Roll, R. (1992). A Mean/Variance Analysis of Tracking Error. Journal of Portfolio Management, 18, 13–22.

Shin S. and Soydemir G. (2010). Exchange-traded Funds, Persistence in Tracking Errors and Information Dissemination. Journal of Multinational Financial Management, 20, 214–234.