A New System Of Choosing The Optimal ETF

September 10, 2023
Jingxiang ZOU

Summary

In the report, we adopt multiple measures to evaluate tracking errors based on recent data and evaluate fund liquidity as represented by volume and a special gauge of bid-ask spread proposed by Lyxor Asset Management. Based on tracking error metrics and liquidity gauges, we propose an ETF efficiency indicator that can be used to choose the optimal ETF for trading purposes.

Integrated Efficiency Indicator

The **Integrated Efficiency Indicator** is a measurement of ETF efficiency that takes into consideration two prime elements of ETF performance: tracking error and liquidity. It uses five widely recognized gauges for tracking errors in academia and industry while utilizing volume and Lynor liquidity spread as proxies for liquidity. The IEI Score for the i^{th} ETF following a certain index can be represented as follows:

$$\mathrm{IEI}_i = \frac{RI(MAE_i) + RI(RMSE_i) + RI(ORC_i) + RI(RDV_i)}{10} + \frac{\mathrm{RI}(\mathrm{Volume_i}) + \mathrm{RI}(\mathrm{Spread_i})}{4}$$

Notes:

1. 'RI' represents the ranking of the ith ETF per a certain measurement, i.e. if $RI(MAE_i) = 5$, it means that the ith ETF corresponds to the 5^{th} lowest mean average error among all candidate ETFs.
2. The denominators are 10 and 2 respectively as to assign equal weights to both elements of ETF efficiency: tracking errors and liquidity.

Empirical Results

Optimal ETF choice for four of the Chinese Equity market indexes (based on minute-wise data from the latest one month):

Index	Best ETF	Sec Code	IEI Score	Lyxor Spread
科创 50	易方达	588080.SH	1.8	15
沪深 300	华夏	510330.SH	1.5	5
中证 500	嘉实	159922.SZ	1.6	7
中证 1000	华夏	159845.SZ	1.8	6

Notes:

1. ‘Optimal Choice’ reports the name abbr. of the fund company that manages the best ETF.
2. ‘Liquidity Spread’ is reported in bp(s).
3. All results are based on the most recent data per 9am, Mon Sep 11th 2023 (UTC+8).
4. All results are based on the given target volume of 300000.
5. About time frame: tracking error gauges are calculated based on minute-wise data from the latest 22 trading days (including today if today is also a trading day) while the Lyxor spreads are computed based on tick data from the latest 7 trading days (including today if today is also a trading day). There is a trade-off between data availability and alignment of time frame and our approach has utilized as many available data as possible.
6. In my code, the results based on 10min-wise data from the latest 3 months are also computed. These data, should they be utilized in some way in the future, are nevertheless not adopted in this report because liquidity can change swiftly making 3 months too wide a time frame.
7. The following chart shows the results when 10min-wise data from the latest 3 months are utilized for the calculation of tracking error gauges and all else the same. The results agree with the one-month chart except for the best choice for the CSI 300 index.

Index	Best ETF	Sec Code	IEI Score	Lyxor Spread
科创 50	易方达	588080.SH	1.8	15
沪深 300	华泰柏瑞	510300.SH	1.4	5
中证 500	嘉实	159922.SZ	1.5	7
中证 1000	华夏	159845.SZ	1.4	6

Lyxor Liquidity Spread

Liquidity is a crucial factor of hedge trading risk management as it substantially affects tracking errors and determines impact cost. Volume and bid-ask spread can both serve as good indicators of liquidity and as suggested by empirical results, they almost always agree on the rankings of candidate ETFs.

This report utilizes the ‘liquidity gauge’ created by Marlène Hassine and Thierry Roncalli from Lyxor Asset Management which was first proposed in their paper in 2013. Unlike the usual bid-ask spread which only considers the highest bid and lowest ask, the Lyxor liquidity spread takes into consideration multiple limit orders in the book and fits well to the scenario when the order size is so large that the final strike price is very different from the market price.

In a nutshell, Lyxor liquidity spread is the discrepancy between the expected strike price while selling and buying a certain worth of asset given the present limit order book. Its rationale can be concisely explained as follows.

Formula (1) is the definition of the Lyxor Liquidity Spread.

$$spread_j = c\frac{\overline{P}_j^+ - \overline{P}_j^-}{\overline{P}_j^0} \tag{1}$$

where

$$\overline{P}_j^* = \frac{\sum_{k=1}^{K} \overline{Q}_{j,k}^* P_{j,k}^*}{\sum_{k=1}^{K} \overline{Q}_{j,k}^*} \tag{2}$$

In formula (2) * stands for either '+'or '-'. $\overline{P}_j^+$ and $\overline{P}_j^-$ stands for the kth limit order ask price and bid price at instant t_j respectively. The average mid-price $\overline{P}_j^0$ in formula (1) is the average of $\overline{P}_j^+$ and $\overline{P}_j^-$.

$$\overline{P}_j^0 = \frac{\overline{P}_j^+ + \overline{P}_j^-}{2} \tag{3}$$

The quantity $\overline{Q}_{j,k}^+$ and $\overline{Q}_{j,k}^+$ are defined as follows.

$$\overline{Q}_{j,k}^* = max\left(0, min\left(Q_{j,k}^*, Q_j^* - \sum_{l=1}^{k-1} Q_{j,l}^*\right)\right) \tag{4}$$

where $Q_{j,k}^+ and\ Q_{j,k}^+$ are the volume of the kth order book at instant t_j. The targeted volume Q_j^* in (4) can be represented by targeted turn volume divided by average mid price as in (5).

$$Q_j^* = \frac{N}{P_j^0} \tag{5}$$

The multiplier c is introduced to accomadate the scenario where order volume is too big for the current book. The multiplier is bigger that 1 in these searios and equals 1 otherwise.

$$c_j = max\left(1, \frac{Q_j^*}{min\left(\sum_{k=1}^{K} Q_{j,k}^+, \sum_{k=1}^{K} Q_{j,k}^-\right)}\right) \tag{6}$$

Tracking Error Metrics

Mean Absolute Error (MAE): The arithmetic mean of the absolute difference between the in-day return of an ETF and that of its corresponding index

$$MAE = \frac{\sum_{t=0}^{T} \left|R_t^{etf} - R_t^{index}\right|}{T - 1}$$

Root Mean Square Error (RMSE): The quadratic mean of the absolute difference between the in-day return of an ETF and that of its corresponding index

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=0}^{T}\left(R_t^{etf} - R_t^{index}\right)^2}{T}}$$

OLS Regression Coefficient (ORC): OLS regression coefficient as per ETF in-day returns regressed on that of the corresponding index.

$$ORC = \left|1 - \frac{\mathrm{Cov}\left(R^{etf}, R^{index}\right)}{\mathrm{Var}\left(R^{index}\right)}\right|$$

Return Discrepancy Volatility (RDV): Standard deviation of the difference between the in-day return of an ETF and that of its corresponding index.

$$\mathrm{RDV} = \sqrt{\mathrm{Var}(R^{etf} - R^{index})}$$

Maximum Absolute Discrepancy (MAD): The maximum among absolute values of in-day return discrepancies.

$$MAD = MAX\left|R^{etf} - R^{index}\right|$$

Bibliography

Hassine, M., & Roncalli, T. (2013). Measuring performance of Exchange-Traded funds. *The Journal of Index Investing*, 4(3), 57–85.

Frino, A., Gallagher, D. R., Neubert, A. S., & Oetomo, T. (2004). Index design and implications for index tracking. *The Journal of Portfolio Management*, 30(2), 89–95.

Rompotis, G. G. (2011). Predictable patterns in ETFs' return and tracking error. *Studies in Economics and Finance*, *28*(1), 14-35.

Rudolf, M., Wolter, H. J., & Zimmermann, H. (1999). A linear model for tracking error minimization. *Journal of Banking & Finance*, *23*(1), 85-103.

Bae, K., & Kim, D. (2020). Liquidity risk and exchange-traded fund returns, variances, and tracking errors. *Journal of Financial Economics*, *138*(1), 222-253.

Appendix A

This appendix contains values of the tracking error metrics and their corresponding rankings.

科创 50					
Fund	MAE	RMSE	ORC	RDV	MAD
华夏	0.0018	0.0027	0.0442	0.0026	0.0128
易方达	0.0013	0.0016	0.0074	0.0014	0.0062
工银	0.0011	0.0016	0.0258	0.0014	0.0065

华泰柏瑞	0.0013	0.0017	0.0318	0.0015	0.0067
南方	0.0019	0.0024	0.0305	0.0023	0.0109
Ranking					
华夏	4	5	5	5	5
易方达	2	2	1	2	1
工银	1	1	2	1	2
华泰柏瑞	3	3	4	3	3
南方	5	4	3	4	4

沪深 300					
Fund	MAE	RMSE	ORC	RDV	MAD
华泰柏瑞	0.00075	0.00103	0.06895	0.00102	0.00506
华夏	0.00070	0.00091	0.02131	0.00086	0.00298
易方达	0.00098	0.00135	0.08282	0.00135	0.01489
天弘	0.00122	0.00166	0.03478	0.00164	0.01766
南方	0.00091	0.00119	0.04288	0.00116	0.01314
Rankings					
华泰柏瑞	2	2	4	2	2
华夏	1	1	1	1	1
易方达	4	4	5	4	4
天弘	5	5	2	5	5
南方	3	3	3	3	3

中证 500					
Fund	MAE	RMSE	ORC	RDV	MAD
南方	0.00078	0.00095	0.02092	0.00091	0.00279
嘉实	0.00068	0.00087	0.00456	0.00087	0.00318
华夏	0.00073	0.00091	0.01655	0.00091	0.00328
广发	0.00110	0.00133	0.07871	0.00128	0.00520
易方达	0.00126	0.00169	0.08797	0.00165	0.01427
华泰柏瑞	0.00164	0.00208	0.07955	0.00207	0.00805
Rankings					
南方	3	3	3	3	1
嘉实	1	1	1	1	2
华夏	2	2	2	2	3
广发	4	4	4	4	4
易方达	5	5	6	5	6
华泰柏瑞	6	6	5	6	5

中证 1000

Fund	MAE	RMSE	ORC	RDV	MAD
华夏	0.00093	0.00120	0.00910	0.00120	0.00388
南方	0.00090	0.00110	0.01216	0.00109	0.00386
富国	0.00087	0.00110	0.04116	0.00108	0.00407
易方达	0.00104	0.00130	0.00703	0.00129	0.00420
华泰柏瑞	0.00154	0.00245	0.12507	0.00245	0.01792
Rankings					
华夏	3	3	2	3	2
南方	2	2	3	2	1
富国	1	1	4	1	3
易方达	4	4	1	4	4
华泰柏瑞	5	5	5	5	5

Appendix B

This appendix contains values of the liquidity metrics and rankings as well as IEI scores.

科创 50					
	Volume	Spread	Volume RI	Spread RI	IEI Score
华夏	19062357	15	1	1	2.9
易方达	4265941	16	2	2	1.8
工银	1658095	16	3	3	2.2
华泰柏瑞	752818	17	4	4	3.6
南方	84575	29	5	5	4.5

沪深 300					
	Volume	Spread	Volume RI	Spread RI	IEI Score
华泰柏瑞	18484411	4	1	1	1.7
华夏	4050480	5	2	2	1.5
易方达	2299387	9	3	3	3.6
天弘	57527	22	4	4	4.2
南方	37565	26	5	5	4

中证 500					
	Volume	Spread	Volume RI	Spread RI	IEI Score
南方	8517415	5	1	1	1.8
嘉实	2619718	7	2	2	1.6
华夏	1309342	7	3	3	2.6
广发	364138	16	4	4	4

易方达	74358	22	5	5	5.2
华泰柏瑞	42315	25	6	6	5.8

中证 1000					
	Volume	Spread	Volume RI	Spread RI	IEI Score
华夏	7688570	6	1	1	1.8
南方	5716582	7	2	2	2
富国	4200232	8	3	4	2.75
易方达	4079315	7	4	3	3.45
华泰柏瑞	90480	68	5	5	5

Appendix C

This appendix contains three empirical probabilities that:

1. The ETF has higher in-day return increment than that of its corresponding index given the index return is positive. (Probability A)
2. The ETF has higher in-day return decrease than that of its corresponding index given the index return is negative. (Probability B)
3. The ETF has either has higher in-day return increment than that of its corresponding index when the index return is positive or higher in-day return decrease than that of its corresponding index when the index return is negative. (Probability C)

科创 50			
	Probability A	Probability B	Probability C
华夏	0.32	0.58	0.47
易方达	0.31	0.67	0.52
工银	0.31	0.68	0.53
华泰柏瑞	0.19	0.68	0.48
南方	0.29	0.58	0.46

沪深 300			
	Probability A	Probability B	Probability C
华泰柏瑞	0.20	0.18	0.19
华夏	0.53	0.33	0.40
易方达	0.27	0.30	0.29
天弘	0.25	0.39	0.34
南方	0.20	0.43	0.35

中证 500

	Probability A	Probability B	Probability C
南方	0.45	0.28	0.35
嘉实	0.42	0.44	0.43
华夏	0.42	0.38	0.40
广发	0.41	0.21	0.29
易方达	0.22	0.38	0.32
华泰柏瑞	0.29	0.39	0.35

中证 1000			
	Probability A	Probability B	Probability C
华夏	0.49	0.56	0.53
南方	0.48	0.40	0.43
富国	0.29	0.44	0.38
易方达	0.45	0.39	0.42
华泰柏瑞	0.47	0.36	0.41