Internet Appendix to "The Smart Beta Mirage"

by Shiyang Huang, Yang Song, and Hong Xiang

Internet Appendix A. Additional Results

This section reports additional results and robustness checks. We conduct several robustness tests for the results about post-listing performance decline: (*i*) We show that the degree of post-listing performance decline is even stronger when we compare index performance before and after the index release date rather than the ETF listing date (see Internet Appendix Table A.1); (*ii*) We show that the results in Table 2 are robust when we require ETFs to have a longer history (at least three years) of non-missing index returns before ETF listings (see Internet Appendix Table A.2); (*iii*) When we loosen our sample restriction and require the sample smart beta ETFs to have non-missing index returns in the three-month window before and after ETF listing, the results are robust (See Internet Appendix Table A.3); (*iv*) We perform a regression analysis for the results in Table 5 and find consistent results (see Table A.4); (*v*) We analyze the influence of decreasing returns to scale based on the weighted-average liquidity of ETF holdings. We find no significant relationship between ETFs' portfolio liquidity and post-listing performance decline (see Internet Appendix Table A.5); (*vi*) We perform a regression analysis for the publication effect in Table 7 and find consistent results (see Internet Appendix Table A.6); (*vii*) We lag the publication date by three years and reperform the analysis for the publication effect in Table 7 (see Internet Appendix Table A.7). (*viii*) We perform a regression analysis for results in Table 11 and find consistent results (see Internet Appendix Table A.13);

We explore another test for the data mining explanation of post-listing performance decline. We hypothesize that the discretion of data mining is larger when the volatility of the underlying index is larger. To test this hypothesis, we classify ETFs into a "higher volatility" and a "lower volatility" group based on the pre-ETF-listing return volatility of

the underlying smart beta index, and we compare the degree of post-listing performance decline between these two groups. Internet Appendix Table A.8 shows that the degree of post-listing performance decline is larger for indexes with higher volatility in the pre-listing period, suggesting that higher volatility is associated with larger discretion of data mining.

In addition, we perform additional analyses for the results of smart beta ETF flows. First, in Internet Appendix Table A.9, we show that post-listing ETF flows respond positively and significantly to the pre-listing index returns in excess of "similar ETFs," where the similar ETFs are those listed ETFs under the same factor theme category of a given ETF. Second, in Internet Appendix Table A.10, we show that, during the $6th$ to $12th$ months after ETF listing, ETF flows exhibit a weak response to the backtested performance but a significantly positive response to the live ETF performance. Third, we analyze the relationship between ETF flows and future ETF performance. Either through portfolio exercise or through panel regressions, we find that investment flows significantly and negatively predict the future performance of smart beta ETFs (see Internet Appendix Table A.11).

Lastly, we conduct a robustness check for Table 10 by using both pre-listing and post-listing index returns, and the results are similar (see Internet Appendix Table A.12).

Table A.1 **Smart beta index performance before and after index release.** This table reports index performance before and after the index release date rather than the ETF listing date. The analysis follows Table **??**. The sample period ends in December 2019. Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Annualized Index CAPM alpha			
	Before	After	Diff
All years before and after index release	3.49%***	−0.41%	−3.90%***
	(9.58)	(−1.28)	(−8.65)
(−1 Year, +1 Year) around index release	1.79%***	0.29%	−1.50%***
	(6.60)	(0.55)	(−2.69)
(−2 Year, +2 Year) around index release	2.26%***	−0.22%	−2.48%***
	(6.97)	(−0.44)	(−4.65)
(−3 Year, +3 Year) around index release	2.34%***	−0.16%	−2.50%***
	(9.76)	(−0.36)	(−5.78)

Table A.2 **Robustness check of Table 2: Require non-missing index returns in three years before listing.** In this table, we re-perform the analysis in Table 2 with the subsample of smart beta indexes that have non-missing returns over at least three years before ETF listing. Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Annualized Index CAPM Alpha				
	#ETF	Before	After	Diff
All years before and after listing	223	2.91%***	−0.43%	−3.34%***
		(4.94)	(−1.63)	(−4.88)
(−1 Year, +1 Year) around listing	223	1.27%***	0.71%	−0.56%
		(4.76)	(1.30)	(−1.05)
(−2 Year, +2 Year) around listing	223	1.21%***	0.33%	−0.88%
		(2.74)	(0.88)	(−1.28)
(−3 Year, +3 Year) around listing	223	1.57%***	0.16%	−1.41%***
		(4.28)	(0.42)	(−2.64)

Table A.3 **Robustness Check of Table 2: Require non-missing index returns in three months around listing.** In this table, we re-perform the analysis in Table 2 with the subsample of smart beta indexes that have non-missing returns in three months before and after ETF listing. Standard errors are clustered by factor theme as categorized by Morningstar, and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Annualized Index CAPM Alpha				
	#ETF	Before	After	Diff
All years before and after listing	254	2.72%***	−0.63%**	−3.17%***
		(4.68)	(−1.96)	(−4.57)
(−1 Year, +1 Year) around listing	254	1.17%***	0.47%	−0.54%
		(4.00)	(0.80)	(−1.05)
(−2 Year, +2 Year) around listing	254	1.14%***	0.14%	−0.84%
		(2.66)	(0.33)	(−1.23)
(−3 Year, +3 Year) around listing	254	1.47%***	−0.03%	−1.33%**
		(4.00)	(−0.06)	(−2.44)

Table A.4 **Robustness Check: Panel Regression Analysis of Table 5.** This table performs panel regression analysis of Table 5. The unit of observation in this analysis is ETF/index-month. The dependent variables in this table are monthly returns/alphas of the smart beta ETF/index. The key independent variable (POST_LISTING) is a dummy variable that equals one for post-ETF-listing observations and equals zero otherwise. In columns (1)-(2), the dependent variable is monthly index returns in excess of the factor theme benchmark index. In columns (3)-(4), the dependent variable is monthly index alpha relative to the factor theme benchmark index. In columns (5)-(6), the dependent variable is monthly index alpha relative to the corresponding academic factor. ETF fixed effects are included in columns (2), (4), and (6). Standard errors are double clustered by time and by ETF. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
DepVar Measure:	Excess Ret relative to Benchmark Index		Alpha relative to Benchmark Index		Alpha relative to Academic Factor	
POST_LISTING	-0.25^{***}	-0.26^{***}	-0.28^{***}	-0.29^{***}	-0.24^{***}	-0.23^{***}
	(-3.25)	(-2.74)	(-3.34)	(-2.71)	(-3.78)	(-2.91)
ETF FE	N	Y	N	Y	N	Y
No. Obs.	52,117	52,117	51,917	51,917	52,117	52,117
Adj. R^2	0.003	0.005	0.004	0.007	0.003	0.008

Table A.5 **Analyzing the influence of decreasing returns to scale: based on portfolio liquidity.** This table reports the relationship between smart beta index performance and ETF portfolio liquidity. We measure the stock-level illiquidity following Amihud (2002) and then compute the portfolio weighted-average illiquidity for each smart beta ETF. We take the negative of the portfolio-level illiquidity as the portfolio liquidity measure. In Panel A, we divide all smart beta indexes into two groups by their ETFs' average portfolio liquidity over the post-listing period. In Panel B, within each factor theme category, we divide smart beta indexes into two groups by their ETFs' average portfolio liquidity. In both panels, columns (1) and (2) show the average annualized CAPM alpha before and after ETF listing across all indexes. Column (3) shows the average after-minus-before-listing difference in index alphas. Column (4) shows the difference in the average after-minus-before-listing index alphas between the two groups. Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Portfolio Liquidity across All ETFs

	(1) Before	(2) After	(3) Diff	(4) Diff-in-Diff
below median	2.64%***	−0.76%**	−3.40%***	0.42%
	(3.34)	(−2.01)	(−3.31)	(0.55)
above median	2.90%***	−0.08%	−2.98%***	
	(5.76)	(−0.21)	(−6.06)	

Panel B: Portfolio Liquidity within Factor Theme Category

	(1) Before	(2) After	(3) Diff	(4) Diff-in-Diff
below median	2.65%***	−0.82%**	−3.47%***	0.59%
	(3.36)	(−2.41)	(−3.39)	(0.69)
above median	2.90%***	0.01%	−2.88%***	
	(5.87)	(0.03)	(−6.22)	

Table A.6 **Robustness Check: Panel Regression Analysis of Table 7.** This table performs panel regression analysis of Table 7. The key independent variable (POST_LISTING) is a dummy variable that equals one in post-listing period and equals zero otherwise. The dependent variables in columns (1)-(5) are smart beta index performance measured by: (1) CAPM Alpha, (2) Alpha relative to SPY, (3) Return in excess of factor theme benchmark index, (4) Alpha relative to factor theme benchmark index, and (5) Alpha relative to academic factors, respectively. To rule out the publication effect, we only use the index performance in the post-publication period. ETF fixed effects are included. Standard errors are double clustered by ETF and time. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

	(1)	(2)	(3)	(4)	(5)
POST_LISTING	-0.27^{***}	-0.31^{***}	-0.25^{***}	-0.28^{***}	-0.23^{***}
	(-2.98)	(-2.78)	(-2.60)	(-2.59)	(-2.86)
ETF FE	Y	Y	Y	Y	Y
No. Obs.	47,053	47,053	47,053	47,011	47,053
Adj. R^2	0.006	0.006	0.004	0.006	0.007

Table A.7 **Robustness Check: Lag Publication Date by Three Years.** This table performs a robustness check for Table 7. In this analysis, we lag the identified publication date by three years, and we only use the index performance in the lagged post-publication period. Panel A reports the annualized index performance in the pre- and post-listing periods. Panel B shows regression analysis on the post-listing performance decline. The key independent variable (Post Listing) is a dummy variable that equals one in post-listing period and equals zero elsewhere. The dependent variables in columns (1)-(5) are smart beta index performance measured by: (1) CAPM Alpha, (2) Alpha relative to SPY, (3) Return in excess of factor theme benchmark index, (4) Alpha relative to factor theme benchmark index, and (5) Alpha relative to academic factors, respectively. ETF fixed effects are included. Standard errors are double clustered by time and by ETF. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Pre- and Post-listing Index Performance

	Before	After	Diff
CAPM Alpha	2.46%***	−0.46%*	−2.92%***
	(4.49)	(−1.81)	(−4.62)
Alpha relative to SPY	2.79%***	−0.88%***	−3.67%***
	(5.15)	(−3.36)	(−5.62)
Ret in excess of factor theme benchmark	2.46%***	−0.87%	−3.33%***
	(6.38)	(−1.64)	(−4.02)
Alpha relative to factor theme benchmark	2.12%***	−1.29%**	−3.41%***
	(5.63)	(−2.18)	(−3.86)
Alpha relative to Academic Factor	1.94%***	−0.66%***	−2.60%***
	(3.90)	(−2.85)	(−3.85)

Panel B: Panel Regressions

	(1)	(2)	(3)	(4)	(5)
POST_LISTING	−0.26***	−0.31***	−0.24**	−0.28**	−0.22***
	(−2.95)	(−2.73)	(−2.48)	(−2.54)	(−2.81)
ETF FE	Y	Y	Y	Y	Y
No. Obs.	44,445	44,445	44,445	44,445	44,445
Adj. R^2	0.005	0.005	0.003	0.005	0.005

Table A.8 **Index Return Volatility and Post-listing Performance Decline.** This table analyzes the relationship between index return volatility and the degree of post-listing performance decline. We split the smart beta ETFs into two halves based on the pre-listing return volatility of the underlying smart beta index. Panel A compares the average annualized CAPM alpha of the indexes in the higher volatility group with indexes in the lower volatility groups. Panel B performs a regression analysis using ETF-month observations. The dependent variable is the monthly CAPM alpha of the underlying index. The key independent variable is the post-listing period dummy variable, the higher volatility dummy variable, and their interaction term. Standard errors are double clustered by time and by ETF. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Higher Volatility vs. Lower Volatility Indexes

Group	#ETF	Before	After	Diff	Diff-in-Diff
Higher Volatility	119	2.73%***	−1.75%***	−4.48%***	−2.53%***
		(3.93)	(−4.25)	(−4.31)	(−3.82)
Lower Volatility	119	2.81%***	0.86%***	−1.95%***	
		(4.95)	(3.27)	(−3.31)	

Panel B: Regression Analysis

DepVar: Idx CAPM Alpha	(1)	(2)	(3)
	Low Vol	High Vol	Full Sample
POST	−0.19**	−0.36***	−0.19***
	(−2.60)	(−3.95)	(−2.61)
POST×HIGH_VOL			−0.17**
			(−1.98)
HIGH_VOL			0.00
			(0.07)
No. Obs.	23,985	28,132	52,117
Adj. R^2	0.002	0.005	0.004

Table A.9 **Pre-listing index performance and post-listing ETF flows.** This table estimates the relationship between the pre-listing backtested performance and the post-listing investment flows. In both panels, the dependent variable is the average monthly percentage ETF flows over the k-month period (k = 6 or 12) after ETF listing. The independent variables are the annualized benchmark-adjusted returns of the underlying index over the one-year, two-year, three-year, and entire time windows before ETF listing. In Panel A, the pre-listing index performance is measured by the index return in excess of the returns of "similar ETFs," where the similar ETFs are those listed ETFs under the same factor theme category of a given ETF. In Panel B, the pre-listing index performance is measured by the index return in excess of the returns of similar ETFs, scaled by the standard deviation of index returns. In all regressions, ETF-listing-year fixed effects and factor theme fixed effects are included. Standard errors are double clustered by factor theme categorized by Morningstar and by ETF listing year. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: ETF-based Factor Theme Benchmark Adjusted Returns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	k=6				k=12			
ONE_YEAR_BEFORE_LISTING	1.54***				1.33***			
	(2.67)				(3.61)			
TWO_YEARS_BEFORE_LISTING		1.15				1.04*		
		(1.64)				(1.81)		
THREE_YEARS_BEFORE_LISTING			3.66***				2.57***	
			(3.94)				(3.66)	
ENTIRE_PERIOD_BEFORE_LISTING				2.96***				2.16**
				(2.66)				(2.06)
Listing year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Obs.	189	189	189	189	189	189	189	189
Adj. R^2	0.196	0.182	0.215	0.200	0.119	0.099	0.129	0.115

Panel B: ETF-based Factor Theme Benchmark Adjusted Returns, Scaled by SD of Benchmark Return

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	k=6				k=12			
ONE_YEAR_BEFORE_LISTING	0.25**				0.22***			
	(2.29)				(2.91)			
TWO_YEARS_BEFORE_LISTING		0.18				0.17		
		(1.43)				(1.63)		
THREE_YEARS_BEFORE_LISTING			0.54***				0.40***	
			(2.81)				(2.71)	
ENTIRE_PERIOD_BEFORE_LISTING				0.48***				0.37**
				(2.74)				(2.32)
Listing year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Obs.	189	189	189	189	189	189	189	189
Adj. R^2	0.200	0.184	0.222	0.209	0.126	0.103	0.140	0.131

Table A.10 **Pre-listing index performance and post-listing ETF flows.** This table estimates the relationship between the pre-listing backtested performance and the post-listing investment flows. In columns (1)-(2), the dependent variable is the average monthly percentage ETF flows over the first six-month period after ETF listing. In columns (3)-(4), the dependent variable is the average monthly percentage ETF flows over the second six-month period after ETF listing. The independent variables include: (1) annualized index returns/alphas in the three-year window before ETF listing, (2) annualized index returns/alphas in the entire period before ETF listing, and (3) annualized ETF returns/alphas in the first six months after ETF listing (PAST_6M_MktExRet). In Panel A, the independent variables are based on CAPM alphas. In Panel B, the independent variables are based on index returns in excess of the factor theme benchmark. In all regressions, ETF-listing-year fixed effects and factor theme fixed effects are included. Standard errors are double clustered by factor theme categorized by Morningstar and by ETF listing year. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: CAPM Alpha

	(1)	(2)	(3)	(4)
Avg Monthly Flows during:	Post-Listing Month 1-6		Post-Listing Month 7-12	
THREE_YEARS_BEFORE_LISTING	3.02***		0.19*	
	(2.59)		(1.72)	
ENTIRE_PERIOD_BEFORE_LISTING		3.86***		−0.42
		(2.58)		(−1.39)
PAST_6M_MktExRet			1.12***	0.69***
			(4.19)	(3.03)
Listing year FE	Y	Y	Y	Y
Factor theme FE	Y	Y	Y	Y
No. Obs.	209	209	209	209
Adj. R2.	0.221	0.231	0.015	0.016

Panel B: Index Ret in Excess of Factor Theme Benchmark Index

	(1)	(2)	(3)	(4)
Avg Monthly Flows during:	Post-Listing Month 1-6		Post-Listing Month 7-12	
THREE_YEARS_BEFORE_LISTING	4.24***		0.97	
	(3.66)		(1.39)	
ENTIRE_PERIOD_BEFORE_LISTING		3.64***		−0.03
		(3.01)		(−0.09)
PAST_6M_MktExRet			1.23***	0.95*
			(2.74)	(1.74)
Listing year FE	Y	Y	Y	Y
Factor theme FE	Y	Y	Y	Y
No. Obs.	209	209	209	209
Adj. R^2	0.224	0.229	0.020	0.015

Table A.11 **ETF flows and future performance.** This table analyzes the relationship between ETF flows and future ETF performance. Panel A reports results from portfolio-sorting exercises. At each month-end, we sort smart beta ETFs into quintiles based on their average flows over the past 6 or 12 months. We hold the ETF portfolios in the next month and compute the AUM-weighted returns. We report average monthly excess returns, CAPM alpha, and FFC4 alpha (in percent) during the holding period of January 2007 to December 2019. Standard errors are calculated using the Newey-West correction of 12 lags. Panel B reports the results from panel regressions. The dependent variable is the ETF return in month t+1. The key independent variable is the average monthly percentage flow over the past 6 or 12 months. We include the past 6-/12-month ETF returns and the ETF AUM at the end of month t as control variables. ETF and time fixed effects are included. Standard errors are clustered by time. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: ETF Portfolios Sorted by Past Flow

Sort by Past 6-Month Flow			
Portfolio	Excess	CAPM	FFC4
1 (Low Flow)	0.68*	−0.10	−0.05
	(1.65)	(−1.24)	(−0.69)
5 (High Flow)	0.46	−0.31***	−0.26***
	(1.10)	(−3.70)	(−2.97)
5−1	−0.23**	−0.21*	−0.21**
	(−2.44)	(−1.89)	(−2.16)
Sort by Past 12-Month Flow			
Portfolio	Excess	CAPM	FFC4
1 (Low Flow)	0.81*	0.04	0.08
	(1.94)	(0.54)	(1.02)
5 (High Flow)	0.54	−0.23**	−0.20*
	(1.28)	(−2.22)	(−1.90)
5−1	−0.27***	−0.27**	−0.29**
	(−2.69)	(−2.15)	(−2.40)

Panel B: Panel Regressions

	(1)	(2)	(3)	(4)
	k=6		*k=12*	
PAST_K_MONTH_FLOW	−0.02***	−0.02***	−0.01**	−0.01**
	(−2.62)	(−2.66)	(−2.29)	(−2.46)
PAST_K_MONTH_RET		−0.09**		−0.14
		(−2.25)		(−1.45)
AUM		0.00		0.00
		(1.09)		(1.15)
ETF and Time FEs	Yes	Yes	Yes	Yes
No. Obs.	13,863	13,863	13,863	13,863
Adj. R^2	0.586	0.586	0.586	0.586

Table A.12 **Robustness check for Table 10.** For each factor theme in a given month, we form an equal-weighted portfolio of all smart beta indexes whose corresponding ETFs have been listed. We then estimate the portfolio's loading on a set of factors, including market (MKTRF), size (SMB), value (HML), momentum (UMD), quality (QMJ), and volatility (VOL). For indexes in each factor theme category, we report their loadings on the designated factor. For growth indexes, we report their loadings on LMH, which is the negative of the HML factor. Standard errors are computed with Newey-West correction of 12 lags. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Factor Theme Category	First Index	Other Indexes	Diff
Value (HML Factor)	0.279^{***}	0.402^{***}	0.123^{***}
	(11.00)	(12.04)	(2.82)
Growth (LMH Factor)	0.307^{***}	0.152^{***}	-0.155^{***}
	(11.02)	(8.90)	(-4.65)
Momentum (MOM Factor)	0.287^{***}	0.243^{***}	-0.045
	(5.86)	(8.93)	(-0.80)
Quality (QMJ Factor)	0.182^{***}	0.098^{**}	-0.085
	(4.36)	(2.09)	(-1.36)
Risk/Vol (VOL Factor)	0.162^{***}	0.163^{***}	0.001
	(3.53)	(3.42)	(0.01)

Table A.13 **Robustness Check: Panel Regression Analysis of Table 11.** This table performs panel regression analysis of Table 11. The sample consists of the 77 smart beta ETFs which invest in the European, Canadian, Australian, or UK equity market. The unit of observation in this analysis is ETF/index-month. The dependent variables in this table are monthly returns/alphas of the smart beta ETF/index. The key independent variable (Post Listing) is a dummy variable that equals one for post-ETF-listing observations and equals zero elsewhere. In columns (1)-(2), the dependent variable is monthly index returns in excess of regional market index returns. In columns (3)-(4), the dependent variable is monthly index alpha relative to the regional market index. ETF fixed effects are included in columns (2) and (4). Standard errors are double clustered by time and by ETF. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

	(1)	(2)	(3)	(4)
DepVar:	Ret in excess of market index		CAPM alpha	
Post Listing	−0.25***	−0.27***	−0.28***	−0.27***
	(−3.49)	(−3.12)	(−4.19)	(−3.25)
ETF FE	N	Y	N	Y
No. Obs.	16,293	16,293	16,293	16,293
Adj. R^2	0.003	0.003	0.004	0.007

Internet Appendix B.

An Illustration of Index Construction

This section presents the screenshots of the rules in determining the number of constituents of a value index whose ETF is offered by one of the major sponsors. As of June of 2020, this ETF managed more than $50 billion of assets. These screenshots are from the methodology document of the index. We choose not to disclose the name of the index and the identity of the ETF sponsor.

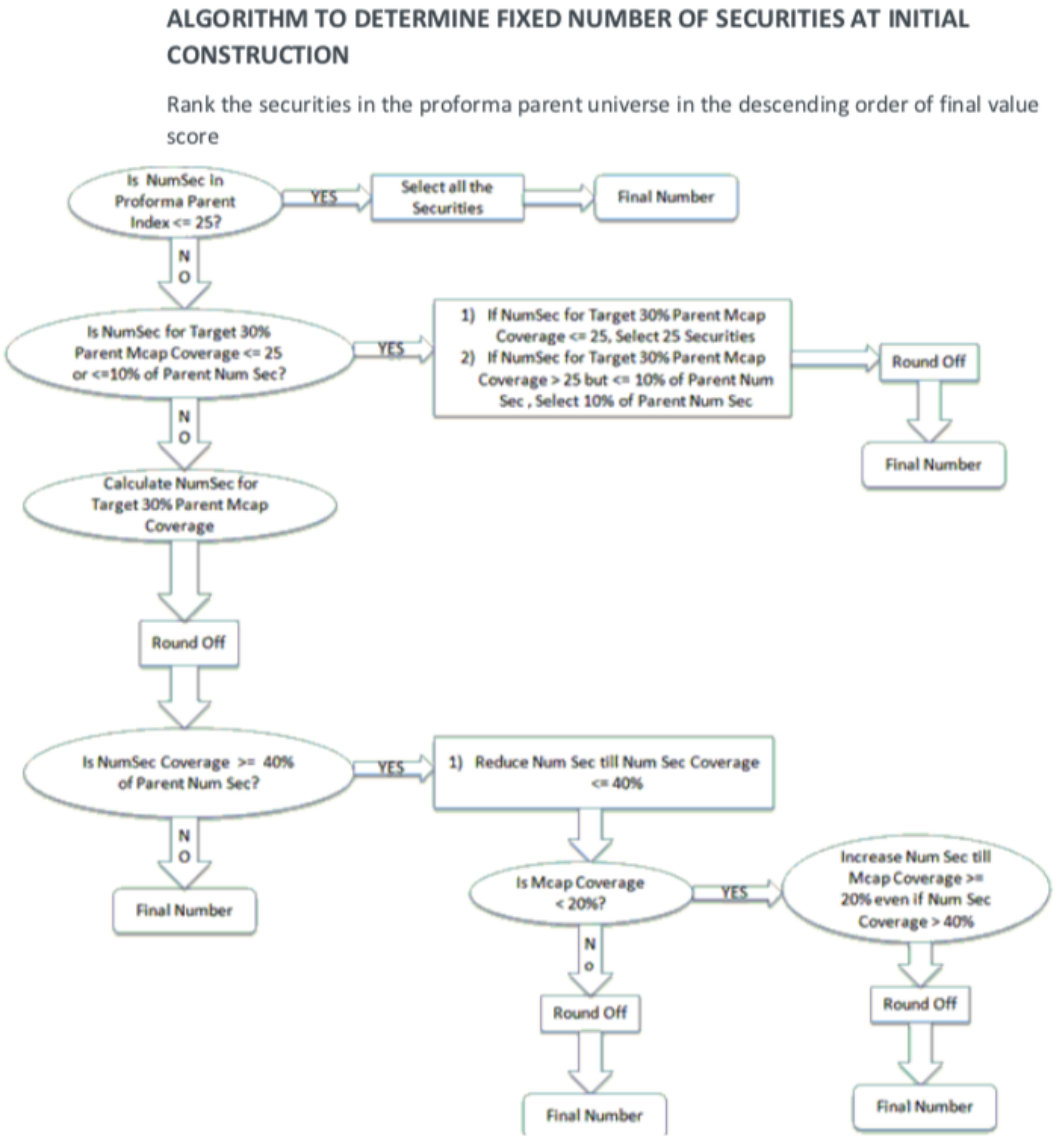

Figure B.1: **Screenshot from the methodology document of a value index.**

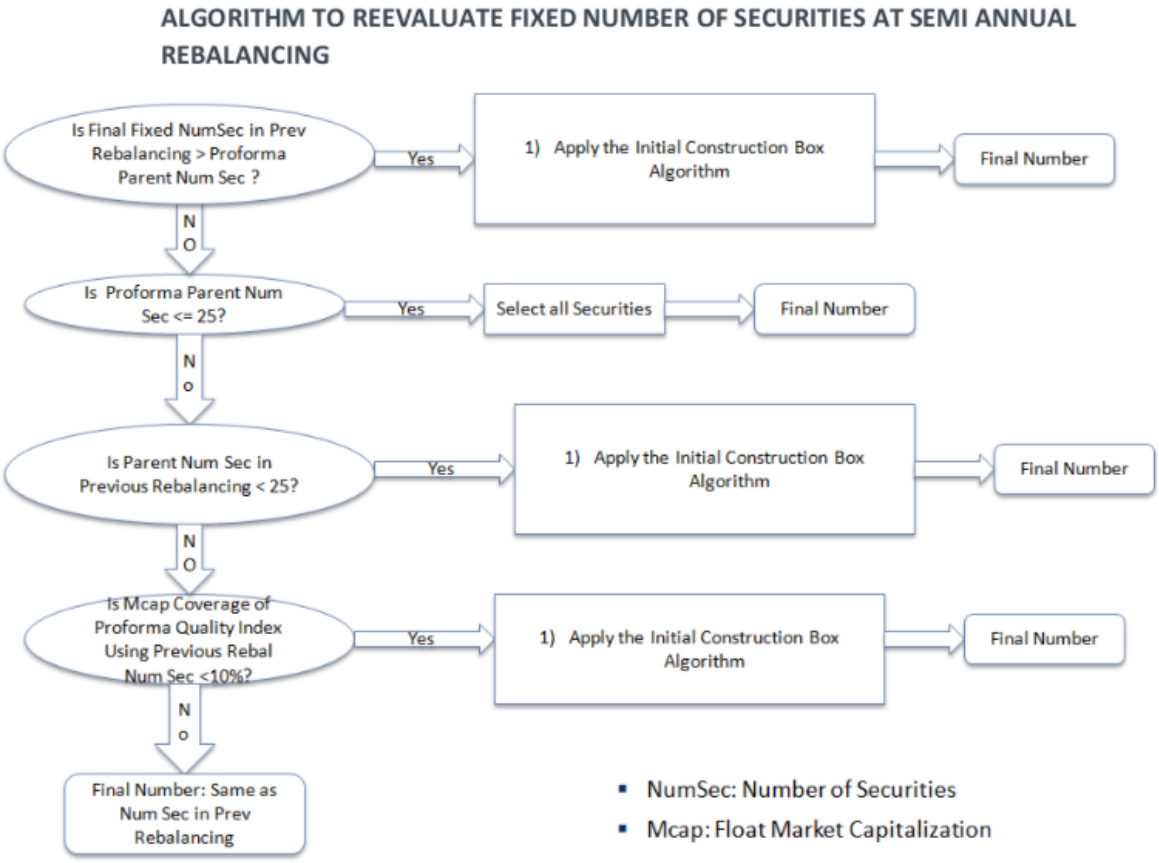

Figure B.2: **Screenshot from the methodology document of a value index.**