Efficacy of Aapryl's Fixed Income Screening Models

In this paper, we will show how Aapryl's return-based systematic screening models can help increase the likelihood of identifying skilled fixed income managers[1], with persistent excess return potential. Analyzing fixed income managers requires an approach which accounts for the primary drivers of their portfolios' performance: interest rate sensitivity (duration), credit risk, and sector differences. Unlike public equity portfolios whose returns are primarily driven by the systematic equity beta factor shared by their benchmarks, fixed income products typically deviate more meaningfully from commonly used benchmarks in critical and multiple dimensions of risk.

The methodologies described herein:

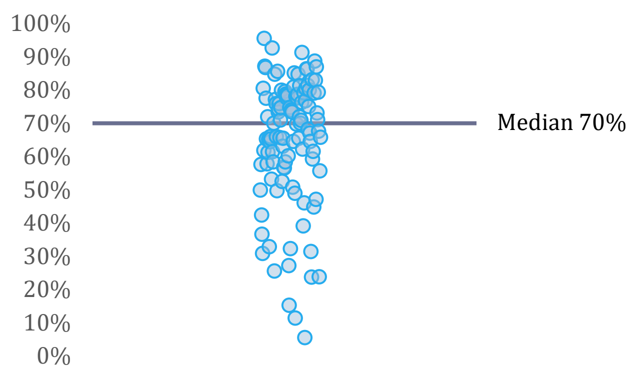
1. Build more accurate benchmarks that account for the deviations from the benchmark with respect to the significant fixed income risks.
2. Build better peer groups that account for salient differences in portfolio characteristics.
3. Build skill measures that leverage the improved benchmarks and peer groups that are explanatory, logical, and predictive.

Problem with Fixed Income Benchmarks

Fixed income benchmarking is an inexact science. **Chart 1** shows the R-squared of all Intermediate Core-Plus Bond funds classified by Morningstar with 10 years of history. The median product only has a 70% R-Squared to its designated benchmark.

Chart 1 Intermediate Core-Plus Bond R-Squared compared to Prospectus Benchmark

(10 years ending 6/30/20)



To put this in context, the Bloomberg Barclays EM Sovereign (USD) index has an R-squared of 74% to the Bloomberg Barclays US HY Corporate Bond index for the past 3 years[2]. While these benchmarks share exposure to broad risk-taking, one would never explain a US High Yield portfolio using an EM Bond index.

[1] The term manager is meant to reflect the decision maker on a given product. It may refer to a person, team or quantitative process.
[2] 36 month R-squared from July 2017 – June 2020 BBgBarc EM USD Sovereign TR USD Index & BBgBarcUS HY Corporate Bonds Index

Although equity benchmarks are imperfect, they are generally more useful for benchmarking public equity strategies. Fully invested equity strategies and their benchmarks are both primarily dominated by equity beta risk. By contrast, the return of fixed income portfolios are driven by multiple dimensions of risk and managers have more flexibility to meaningfully alter their exposure to these return drivers. Principal Component Analysis[3] is a useful technique to test how many independent risk drivers there are within multiple return streams.

Chart 2 Percentage of Explained Variances

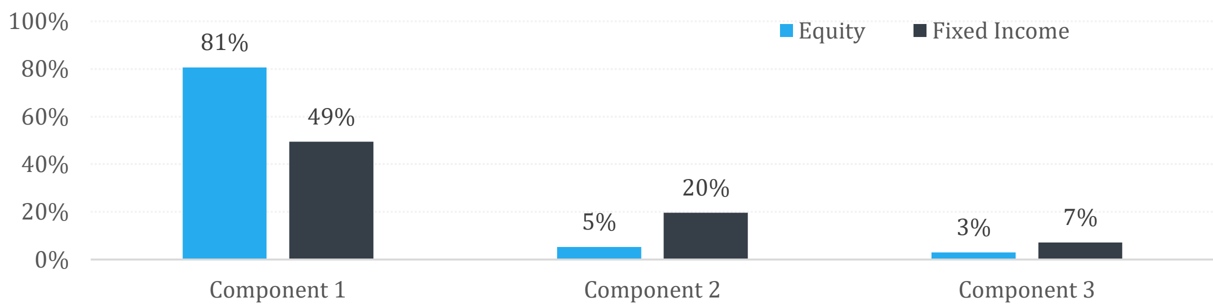
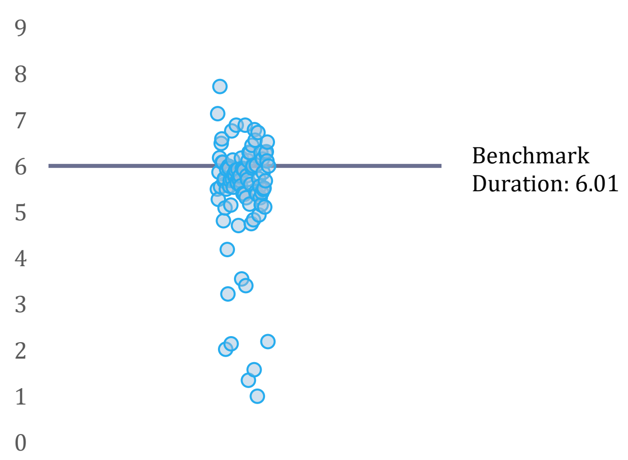


Chart 2 shows that equity indices (including smart beta and factor indices) have a common risk factor that can explain over 80% of their variance. We find it useful to think of this as "equity risk," which all fully invested equity products share in common. Fixed income indices have more diverse risks, with 3 factors failing to explain variance as much as the single equity risk factor.

Further complicating things is the freedom that fixed income managers have to deviate from their benchmark, particularly with respect to duration and spread risks. For example, the average effective duration of the Intermediate Core-Plus Bond peer group ranges from 1.01 to 7.74. (Chart 3) Regardless of how well duration risk explains returns, having such a dispersion among strategies makes comparison impossible. In contrast, U.S. equity managers in particular are typically measured relative to narrowly prescribed benchmarks and are generally penalized for deviating from their benchmark's primary style characteristics.

Chart 3 Intermediate Core-Plus Bond Average Effective Duration
(10 years ending 6/30/20)



[3] PCA run on monthly returns of all factor indices used within the Aapryl system (7/2001 – 12/2020)

One approach for addressing the limitations of market-based fixed income benchmarks is to build custom benchmarks by evaluating the holdings of a manager. While more precise on an individual manager basis, this approach is obviously more labor intensive. But more importantly, this approach would be difficult to scale across a full universe of products, which would be necessary for both manager screening and accurate peer group comparisons.

Building Better Benchmarks through Aapryl's Product Clones

Our model attempts to replicate the spirit of the in-depth approach using returns-based analysis. The result is a replication portfolio which we call the Static Clone. We start by identify exposures that are significant to each manager product through a regularized multistep regression technique that performs both factor selection and beta regularization. We further applied a constrained optimization technique to obtain realistic constraints to the regression outcomes that require all factor exposures to sum to 100%. To customize the duration exposure of our custom benchmarks, we allow for both positive and negative weights to Treasury indices of fixed maturity. Overall gross exposure is capped at 200%. These steps help the regression to avoid risk exposures that are unrealistic to replicate, rendering each manager Static Clone as a transparent and investable replication portfolio that would be straightforward for investors to implement. Below are some summary results showing the improved fit of the product-specific Static Clones relative to traditional benchmarks.

We evaluated the 8,027 fixed income products in the Aapryl system that were not pure index funds and had at least three years of continuous performance. The product count for the strategies that we analyzed are as follows:

Aapryl Category	Product count	Aapryl Category	Product count
Bank Loans	275	HY Munis	158
Core	1,111	MBS	229
Core Plus	398	Municipals	1,947
EM Bonds	420	Short Term Investment Grade	668
Flexible Allocation	475	US Corporate	461
Global Bonds	669	US TIPS	217
High Yield	648	US Treasury	351
		All Fixed Income Products	**8,027**

Tables 1 & 2 show that for all the major fixed income style classifications represented by these products, the Static Clone portfolios have a higher R^2 and a lower tracking error than their respective market benchmarks.

Table 1[4]

	Benchmark R-Squared	Static Clone R-Squared	Change
US Treasury	91.0%	94.2%	3.1%
US TIPS	83.8%	83.8%	0.0%
Core Plus	66.9%	92.7%	25.9%
Core Plus	85.6%	94.8%	9.2%
Municipals	92.9%	95.5%	2.6%
MBS	78.4%	91.6%	13.2%
Bank Loans	98.4%	98.7%	0.3%
High Yield	96.5%	97.0%	0.5%
Flexible Allocation	9.8%	87.8%	78.0%
US Corporate	88.4%	96.2%	7.7%
Short Term Investment Grade	75.7%	90.8%	15.1%
HY Munis	89.3%	92.8%	3.5%
Global Bonds	38.2%	87.2%	49.1%
EM Bonds	88.2%	93.6%	5.4%

Table 2

	Benchmark Tracking Error	Static Clone Tracking Error	Change
US Treasury	1.8%	0.6%	-1.2%
US TIPS	1.5%	1.1%	-0.3%
Core Plus	2.1%	1.1%	-1.0%
Core Plus	1.3%	0.7%	-0.6%
Municipals	1.3%	0.8%	-0.6%
MBS	1.0%	0.7%	-0.3%
Bank Loans	1.1%	0.8%	-0.3%
High Yield	1.6%	1.4%	-0.2%
Flexible Allocation	5.8%	2.2%	-3.5%
US Corporate	1.9%	1.0%	-1.0%
Short Term Investment Grade	1.1%	0.7%	-0.3%
HY Munis	2.5%	1.6%	-0.9%
Global Bonds	4.4%	2.0%	-2.4%
EM Bonds	4.6%	2.9%	-1.6%

In other words, the manager product clones more accurately identify performance that is solely due to significant risk exposures. Not surprisingly, the increase in R-squared and decrease in tracking error are most substantial in categories that have the most significant degrees of freedom for risk factor exposure, such as the Flexible Allocation category.

Building Better Peer Groups

Identifying product specific risk exposures is the first step in improving peer groups. Next, one must group products based on their shared risk profile, rather than the manager's self-identified peer group. Our approach was to analyze each category along the 2 primary risk dimensions; duration and spread risk. To do this, we calculate the implied OAS and Duration of each product's Static Clone. The OAS and Duration of the clone is the weighted average of the long-term average levels of the individual components.

[4] R-Squared and Tracking Error Calculations ITD for all Non-Index Fixed Income products with >36m performance in the Aapryl System (PSN and Lipper fixed income SMAs, Mutual Funds and ETFS, 1Q 203 – 4Q 2020). Aapryl Categories used for aggregation. (Benchmarks used: BBBarc US Treasury, BBBarc TIPs, BBBarc Agg Bond, BBBarc Agg Bond, BBBarc Municipal Bond, BBBarc MBS, BBBarc Agg, CS HY Loans, BBBarc US HY Corp, BBBarc Agg Bond, BBBarc Credit, BBBarc 1-3yr Credit, BBBarc HY Muni, BBBarc Global Agg, BBBarc EM Agg)

Using a K-Means clustering algorithm, we identified logical breakpoints within each category. The theory being that strategies focused on a specific duration / credit risk are best compared with others focusing on the same risk levels.

The different colors in **Chart 4** represent the clusters that the K- Means algorithm recommended for Core Plus category. This analysis provided us with logical breakpoints within each category for peer group construction.

Chart 4 Core Plus Strategy

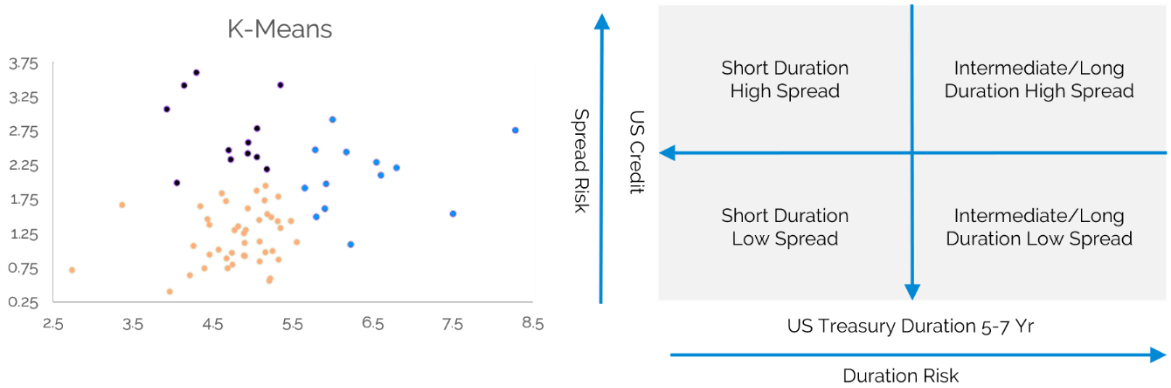
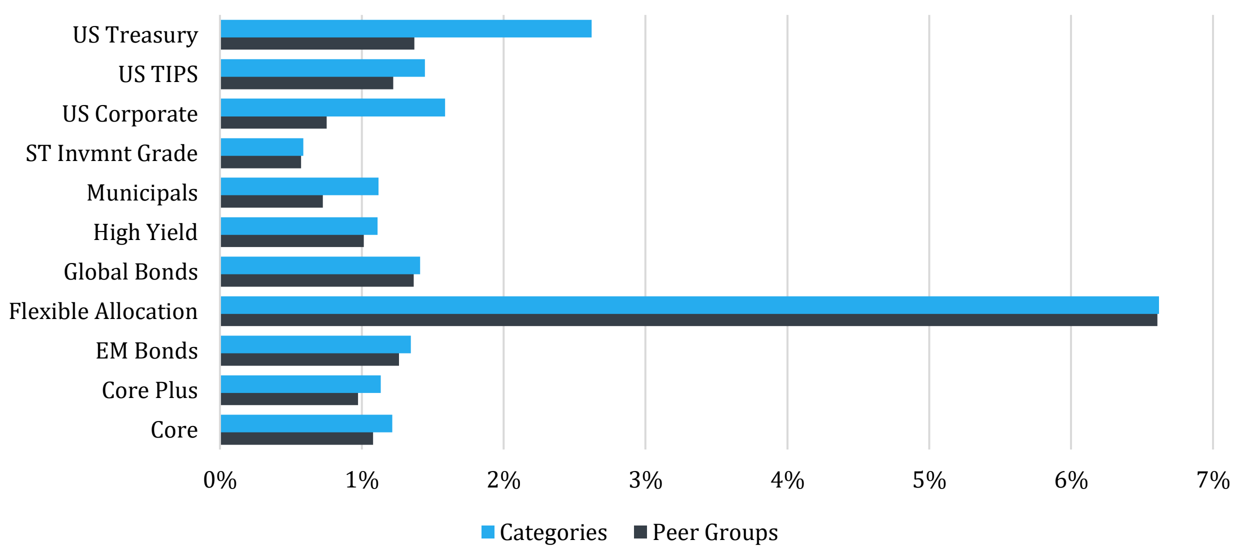


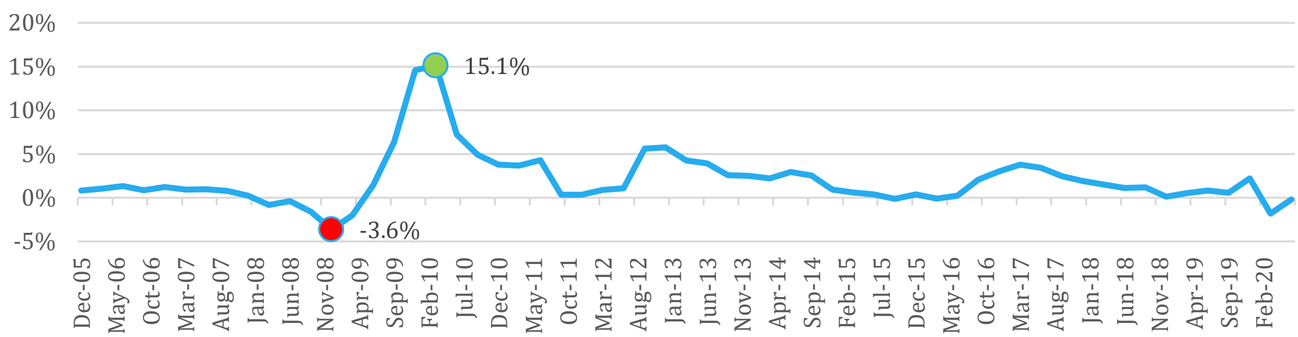
Chart 5 below shows that these risk exposure-based peer groups do a better job of grouping similar managers than traditional classification systems. For all peer groups, the dispersion of returns is smaller when using Aapryl custom peer groups vs the broader categories for classification.

Chart 5 Cross Sectional Standard Deviation of Return Categories vs. Aapryl Peer Groups



Accurate benchmarks and peer groups are essential for manager comparison. To increase the likelihood of finding future outperformance, one must isolate the "skill" of a manager or PM team. Beyond that, one must adjust for the market conditions in which the returns were generated. This is best done by comparing metrics across a well-designed peer group. **Chart 5** shows that the level of Alpha required to be a top quartile manager is very sensitive to the time period and peer group in which it is generated.

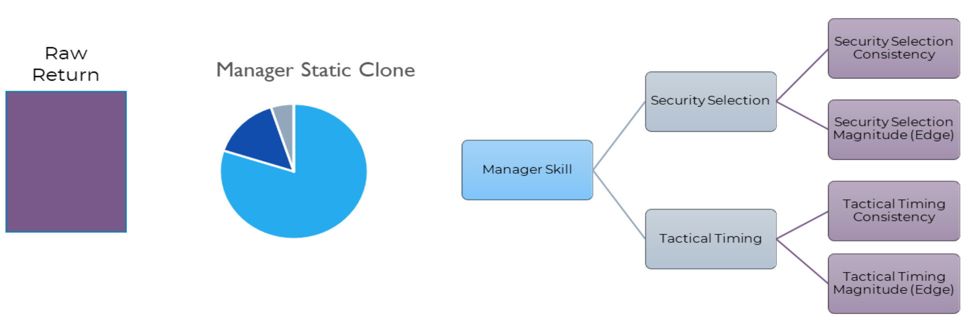
Chart 6 How Much Alpha Would You Need to be a Top Quartile Manager?
Trailing 12 Months (Evestment Core Plus Peer Group)



Building Metrics to Identify Persistent Skill

Our objective is to build metrics that are explanatory, logical, and most importantly, predictive. **Chart 7** below demonstrates how we leverage our Clone methodology and improved peer groups to identify skill. Our methodology distinguishes whether the manager's skill is derived from security selection or tactical timing and further contextualizes the texture of their return. For example, some managers have a clearly defined edge in pricing and execution. They will consistently build their portfolios for a lower cost than their competitors. This skill would likely result in small incremental Alpha over time, which would most likely manifest as consistency. Other managers will identify mispriced sectors or credit opportunities and pursue larger, but less frequent return opportunities. This edge would be better identified through a magnitude measure, which we call edge. The inset on Page 7 describes this methodology in greater detail.

Chart 7



Aapryl Methodology for Contextualizing Manager Skill

To separate the impact of tactical timing from security selection, we expand our use of clone building to a shorter time horizon. The Dynamic Clone utilizes the same methods as the Static Clone, run using a proprietary time-weighting of the trailing 36 months returns of product and factors.

Excess Returns

- *Total Excess Return:* Manager Skill = Product performance minus Static Clone Performance
- *Security Selection Return*: Product performance minus Dynamic Clone Performance
- *Tactical Timing Return:* Dynamic Clone Performance minus Static Clone Performance

Skill Measurements (Calculated for each of the 3 excess return series)

- ***Consistency:*** Measure of the frequency of positive excess return. The foundation is a batting average type of calculation. The final score adjusts for the length of the track record, market conditions and cross-sectional peer group performance.
- ***Edge:*** Measure of the relative magnitude of positive vs. negative excess returns. The measure begins with an Omega ratio (0% threshold) calculation, adjusting for the length of the track record, market conditions, and cross-sectional peer group performance.

Model Performance Summary

To test the efficacy of this methodology, we designed straightforward algorithms that establish critical thresholds for each skill metric based on historical data and tracked how well products that cleared the thresholds performed relative to their peers out of sample.

For example, in our "in sample" analysis, we evaluated all Core plus managers in the Aapryl system[5], through the fourth quarter of 2011 in order to determine what skill metrics and score thresholds helped identify products that had the highest likelihood of finishing in the top-quartile for the following 3 years. The first threshold was the 70th percentile for the Total Consistency metric. The second was the 60th percentile for the Security Selection Consistency metric.

For our "out of sample" analysis, we used the thresholds described above beginning in the first quarter of 2012. We then calculated the following for the 3 years ended 4Q 2014:

- Excess Return vs Static Clone
- Peer Group Percentile ranking

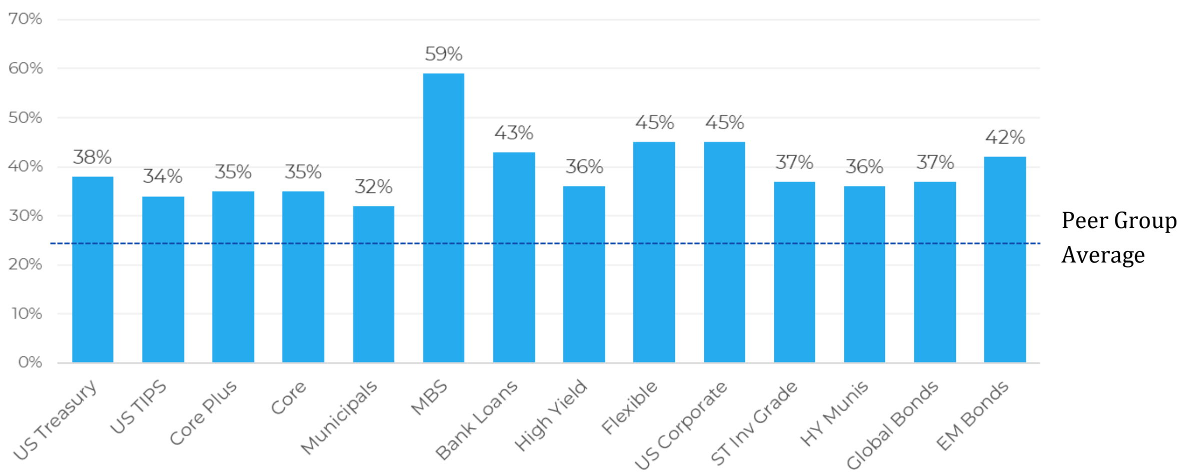
The model was recalibrated every year, with only out of sample data used for efficacy testing purposes. Testing all products in the Aapryl system in this way, the results show the clear predictive quality of the

[5] Excluding all passively managed products PSN Fixed Income SMAs, Lipper Fixed Income Mutual Funds (primary share class), Lipper ETFs. Skill metric calculations begin 1/2003, with 1/2006 being the first score for each metric. The 1st out of sample period begins 1/2009 using the cutoffs from 1/2006 optimized to maximize % of products finishing in the top quartile for the 3 years ended 1/2009.

Aapryl skill metrics. We addressed potential survivorship bias by using the historical database which incorporated dead funds.[6]

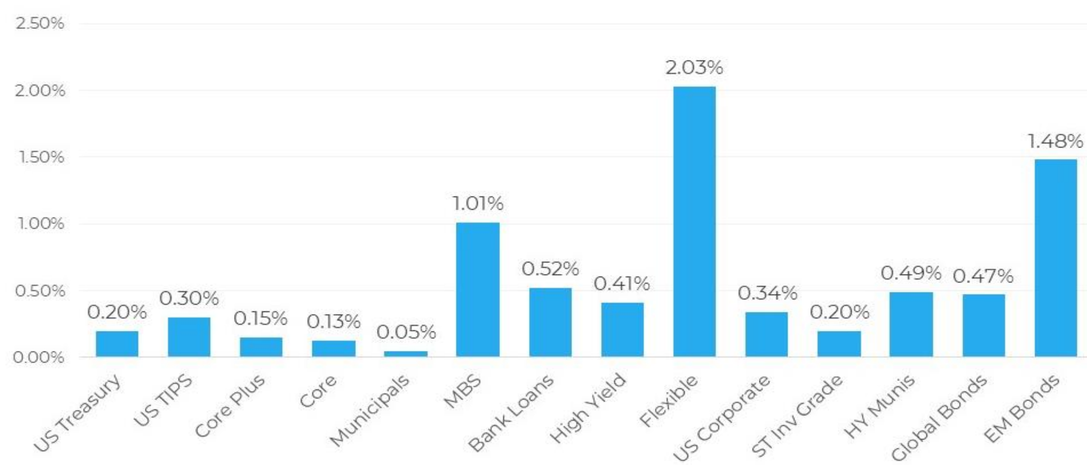
Chart 8 evaluates the efficacy of this model for predicting manager products that have the highest likelihood of placing in the top quartile for the subsequent three years average relative to the random selection level of 25%. Chart 9 evaluates the annualized excess return of fixed income products that cleared our thresholds relative to their peer group averages.

Chart 8[7] Likelihood of Top "Passing" Products Finishing in the Top Quartile
Forward Looking 36 month



For all categories, the products that had skill metrics above the designated threshold had a much greater likelihood of finishing as a top quartile manager in the subsequent 3 years relative to peer group average.

Chart 9 Average Alpha of "Passing Products" vs. Peer Group Average



[6] Another potential bias is that a product needed 6 years of consecutive returns (3 years to generate a score, 3 years of forward-looking returns) to be included in our model. If products that are not included (I.e., less than 6 years) had lower returns than average, then their exclusion could bias returns. To account for this, we analyzed the peer group relative performance of top ranked products. Thus, both the top ranked products and the peer group excluded products with less than 6 years of history.
[7] "Passing" products are all products with scores above the threshold set during in sample testing. Returns are judged the following 3 years, using only out of sample data.

For all categories, the products that had skill metrics above the designated threshold had a much greater annualized excess return (vs their Static Clone) in the subsequent 3 years compared to their peer group average. Not surprisingly, the increase in annualized excess return is greatest for those categories in which standard market benchmarks are least proficient in capturing the multi-dimensional profile of active fixed managers (such as EM bonds or MBS); or not sufficiently dynamic to capture more flexible strategies. Whilst the increase in annual excess return generated by our model might seem modest for the Core or Core Plus categories, it is important to note that the average difference between top quartile and median products over the past 10 years was 48 bps[8] for Core products and 72 bps for Core Plus products. Therefore, the model was able to capture over 25% of the alpha offered by Core Fixed Income products using an entirely systematic approach.

Finally, it is important to note that these model results do not incorporate any qualitative judgements or other manager or product due diligence techniques. The Aapryl skill screening module should be viewed as a **starting point** from which skilled researchers can apply various portfolio analysis and qualitative judgements to further enhance the probability of choosing a skilled manager.

Conclusion

The biases of fixed income benchmarks are a well-established problem. The labor intensive, holdings-based approach to customizing benchmarks is comprehensive, but difficult to scale across the full universe of products. This universal scaling is an important feature of our predictive metrics. To isolate the persistent components of a manager, return stream one must:

1. Build better benchmarks to normalize for product specific risk exposures
2. Build better peer groups to compare products operating with the same opportunity set and risk guideposts
3. Build robust metrics that leverage both the improved benchmark and peer groups in a way that takes the market environment and length of track record into account.

[8] Average of Rolling 3 year annualized returns for all Core & Core Plus Fixed Income Products in the Evestment Universe from 2/2007 - 5/2020 (Core Universe had 484 products, Core Plus had 252).