THE JOURNAL OF
indexinvesting
ETFs, Indexing & Mutual Funds
SUMMER 2012 VOLUME 3 NUMBER 1 | www.IIJII.com

IIJ Institutional Investor Journals
The Voices *of* Influence | iijournals.com

On the Performance of Fixed-Income Exchange-Traded Funds

PATRICK HOUWELING

Robeco is an international asset manager offering an extensive range of active investments, from equities to bonds. Research lies at the heart of everything we do, with a 'pioneering but cautious' approach that has been in our DNA since our foundation in Rotterdam in 1929. We believe strongly in sustainability investing, quantitative techniques and constant innovation.

ROBECO
The Investment Engineers

On the Performance of Fixed-Income Exchange-Traded Funds

PATRICK HOUWELING

PATRICK HOUWELING
is a vice president at Robeco Quantitative Strategies in Rotterdam, The Netherlands.
p.houweling@robeco.com

We evaluate the performance of fixed-income exchange-traded funds (ETFs). ETFs are very similar to open-ended mutual funds but are listed and traded like a single stock on stock exchanges. More and more investors are putting their money in fixed-income ETFs to receive exposure to Treasury bonds, investment-grade corporate bonds, or high-yield corporate bonds. Since the introduction of the first fixed-income ETF in 2000, fixed-income ETFs have exploded in both number and in assets under management (see BlackRock [2010b]).

Because ETFs are specifically designed to track their benchmarks, the foremost measure of an ETF's quality is its ability to replicate benchmark returns. We find that Treasury ETFs have indeed, on average, been able to track their benchmarks. Investment-grade corporate bond ETFs and, especially, high-yield corporate bond ETFs have lagged their benchmarks, however: Their average underperformances are larger than their average costs.

We find that the transaction costs of the underlying bonds are a key determinant of an ETF's underperformance. Because of the mechanics of the creation/redemption process, trading costs find their way into ETF performance. We extract several determinants of transaction costs from the academic literature, such as credit rating, maturity, and size, and create groups of ETFs based on these determinants. We find that ETF underperformance indeed increases with transaction costs, as hypothesized. Therefore, we conclude that an ETF's ability to replicate its benchmark depends on the transaction costs of that benchmark's constituents.

Our results have several implications. First, when forming an expectation about the expected underperformance of an ETF relative to its benchmark, investors should take into account not only stated expense ratios but also the transaction costs of the securities in that benchmark. Especially for high yield, and to a lesser extent for investment grade, investors should expect ETFs to underperform by more than their costs. Consequently, investors should lower their return expectations on corporate bonds in asset allocation decisions when using ETFs in the implementation. Second, our results contribute to the debate on the added value of active management (see, e.g., Derwall and Huij [2008] and Lipton and Kish [2010]). Given that passive corporate bond ETFs cannot replicate benchmark returns, it seems that those benchmark returns are not a fair yardstick for evaluating the performance of active funds either. Instead, by comparing active funds' returns to those of passive funds such as ETFs, investors may more fairly measure these funds' added value.

Our study is similar to Rompotis [2008], Blitz and Huij [2011], and Blitz, Huij, and Swinkels [2012] on equity ETFs and to Drevonak,

Uroševic, and Jelic [2010] on European sovereign bond ETFs. This article is more extensive than Drevonak et al. [2010] because we analyze not only sovereign bond ETFs but also corporate bond ETFs and not only euro ETFs, but also U.S. dollar ETFs; further, our data start in 2002, whereas their data start in 2007.

ETFs[1]

Definition

Exchange-traded funds are very much like open-ended mutual funds, except that they are listed and traded like a single stock on stock exchanges. Traditionally, an ETF is a passive investment fund, aiming to track its benchmark as closely as possible.[2] The first ETF was introduced in 1993: the SPDR (Standard & Poor's Depositary Receipts) on the S&P 500 Index. ETFs are available on many asset classes, including equity, fixed income, currencies, commodities, real estate, private equity, and hedge funds. In this article, we focus on fixed-income ETFs. The first ETF on a fixed-income index was introduced in 2000. At the end of September 2010, that number had increased to 357, managing a total of US$210 billion (BlackRock [2010b]).

Creation and Redemption Process

ETF shares can be created or redeemed only by so-called *authorized participants*. They buy blocks of tens of thousands of ETF shares directly from the ETF provider in so-called *creation units*. Similarly, authorized participants can sell such blocks, in this case called *redemption units*, to the ETF provider. Creations and redemptions are mostly done *in kind*, meaning that an authorized participant exchanges shares for a portfolio of securities held by the ETF. Through this creation/redemption process, authorized participants act as liquidity providers in ETF shares. To the extent that an ETF's underlying securities can be traded in large amounts within short periods and at limited costs (i.e., are sufficiently liquid), the creation/redemption process also acts as an arbitrage mechanism, bringing an ETF's market price in line with the net asset value (NAV) of its underlying securities. Whenever the liquidity in the underlying securities dries up, an authorized participant may be unable to create or redeem ETF shares at the NAV, so that the ETF's market price may start to drift away from the NAV.

Not all fixed-income ETFs use this process, however. Many corporate bond ETFs' prospectuses specify that creations and redemptions occur partially in kind and partially with cash amounts, so-called *in-cash creation*. This approach also limits the arbitrage mechanism and may lead to larger deviations between the ETF's market price and its NAV. Tucker and Laipply [2010] describe the in-kind and in-cash creation methodologies in more detail. They concluded that for ETFs that use in-cash creation, the transaction costs of the underlying securities end up in the ETF's performance, thus affecting all investors in a fund. For ETFs that use in-kind creation, new entrants bear the creation costs.

Tracking Techniques

An ETF provider may use one of three approaches to construct an ETF's portfolio: 1) *full replication*: the provider buys exactly the same bonds and in the same quantity as in the benchmark, 2) *statistical replication*: the provider buys a subset of the bonds in the benchmark aimed at following the benchmark as closely as possible (e.g., using mean–variance optimization), and 3) *swap-based replication*: the provider buys certain securities and in addition engages in a swap contract swapping the return on these securities against the benchmark return.

Ex ante, one would expect better benchmark tracking ability for ETFs that follow full replication or swap-based replication, while the return of ETFs that follow statistical replication may deviate more from the benchmark they track.

HYPOTHESES

In this study, we are interested in the performance of fixed-income ETFs. More specifically, we investigate the difference between an ETF's net total return (including dividends, costs, and fees) and its benchmark's total return. We refer to this return difference as the ETF's *outperformance* (or, if this difference is negative, as its *underperformance*).[3] Further, we define the *cost-adjusted outperformance* as the outperformance plus the costs and fees the ETF withholds. In this section, we derive several hypotheses on the cost-adjusted outperformance from the creation/redemption processes and tracking techniques described previously. We test these hypotheses on empirical data in the Results section.

H1: *For Treasury bond ETFs, the cost-adjusted outperformance is equal to 0.*

As argued by Tucker and Laipply [2010], for ETFs that use the in-kind creation methodology, which Treasury bond ETFs do, the ETF's return should not be affected by inflows or outflows. Moreover, Treasury bonds are among the most liquid instruments in today's financial markets, making them quickly and cheaply tradable in large quantities so that transaction costs are low. So, even if some creation costs drip through to the ETF's return, their impact should be limited. Therefore, we hypothesize that Treasury bond ETFs can track their benchmarks—that is, their net return should lag the benchmark return by exactly the total expense ratio. This implies that the cost-adjusted outperformance should be 0.

H2: *For corporate bond ETFs, the cost-adjusted outperformance is negative.*

Because most corporate bond ETFs use the in-cash creation methodology (sometimes fully, sometimes in combination with in-kind creation), the transaction costs of purchasing the underlying securities (at least partially) end up in the ETF's performance. Given the tremendous growth in assets under management, the potential impact on ETF return may thus have been substantial. Moreover, transaction costs for corporate bonds are larger than for Treasury bonds (e.g., as shown by Chakravarty and Sarkar [2003]), amplifying the potential impact. Therefore, we expect that corporate bond ETFs lag their benchmarks by more than their costs, and we thus expect their cost-adjusted outperformance to be negative.

Not all corporate bonds have the same amount of transaction costs. In addition to Chakravarty and Sarkar [2003], other studies such as Houweling, Mentink, and Vorst [2005] and Chen, Lesmond, and Wei [2007] found various determinants of corporate bond liquidity and transaction costs. Using our data, we can test the impact of three transaction cost determinants on ETFs' performance: credit rating, maturity, and size (see hypotheses H3, H4, and H5).[4]

H3: *The cost-adjusted outperformance of corporate bond ETFs is more negative for high yield than for investment grade.*

Chakravarty and Sarkar [2003] and Chen et al. [2007] showed that bid–ask spreads increase with credit rating—that is, lower-rated bonds are less liquid and more costly to trade. Therefore, we hypothesize high-yield ETFs to underperform more than investment-grade ETFs.

H4: *The cost-adjusted outperformance of corporate bond ETFs is more negative for benchmarks with longer maturity bonds.*

Chakravarty and Sarkar [2003] and Chen et al. [2007] also showed that bid–ask spreads increase with time to maturity. Therefore, we expect ETFs with benchmarks consisting of longer-maturity bonds to underperform more than ETFs on shorter-maturity bonds.

H5: *The cost-adjusted outperformance of corporate bond ETFs is more negative for "broad" benchmarks than for "liquid" benchmarks.*

Houweling et al. [2005], among others, showed that bond liquidity increases with issue size. A subset of the ETFs in our sample has so-called "liquid" benchmarks, which typically consist of bonds with larger outstanding amounts. We hypothesize that ETFs with such benchmarks have smaller underperformances, because the underlying bonds are cheaper to trade. Moreover, these benchmarks contain a smaller number of constituents, making it easier to apply the full replication technique, further increasing the likelihood of generating benchmark-like returns. On the other hand, "non-liquid" or "broad" benchmarks consist of a greater number of less liquid bonds, so using the reverse reasoning, we expect ETFs on broad benchmarks to have bigger underperformances.

In the next section, we introduce the data on which we test these hypotheses.

DATA

We use the overview provided by BlackRock [2010a] to determine which ETFs existed at December 31, 2009. In its quarterly publication, BlackRock, the largest global player in ETFs, provides a complete overview of all existing ETFs. This quarterly overview is widely used in the industry.

We restrict our analyses to "plain vanilla" ETFs with a benchmark consisting of either Treasury bonds, investment-grade credits, investment-grade corporate bonds, or high-yield corporate bonds.[5] Further, we consider benchmarks for only the United States or Europe. We exclude ETFs that provide inverse or leveraged exposure to such benchmarks. We also exclude ETFs with

aggregate or inflation-linked benchmarks. After applying these criteria, we have 129 ETFs. Many ETFs are listed on multiple exchanges. For our analyses, we use only the first listing. For each ETF, we collect the following information from BlackRock [2010a]: ISIN, benchmark index ticker, total expense ratio (TER, as reported by the fund over 2009), and assets under management (AuM, as of June 30, 2010).

For each ETF, we download net daily total returns from Thomson Financial Datastream using its ISIN as identifier. These returns take paid dividends and withheld costs and fees into account. The first Treasury and investment-grade ETFs in our sample start in July 2002. The first high-yield ETF starts in April 2007. All return data end July 31, 2010. Note that we use total returns based on actual market prices, as published by the exchanges, not NAVs published by the ETF providers. Market prices reflect the actual prices at which investors can buy and sell the ETFs. NAVs, on the other hand, are just "paper numbers." As noted earlier, the NAV should closely resemble the market price, but they are not necessarily equal.

For each benchmark, we download daily total returns from Bloomberg, using its ticker as identifier. For a few Barclays Capital indices, return data are unavailable from Bloomberg, and we thus obtained these returns directly from Barclays Capital's website. It is worth noting that the vast majority of the 129 ETFs in our sample have unique benchmarks. In total, the 129 ETFs have 121 benchmarks. Another noteworthy feature of the benchmark indices used by ETF providers, already briefly mentioned, is that 31 of them contain the word "liquid" in their name, including 6 out of 20 investment-grade benchmarks and all three high-yield benchmarks.

Exhibit 1 breaks down the 129 ETFs in our sample by benchmark type (Treasury, investment grade, or high yield). By far, most of the fixed-income ETFs in our sample have a Treasury benchmark—106 out of 129. Next is investment grade, with 20 ETFs. Only three high-yield ETFs existed at the end of 2009. In terms of AuM, the differences are smaller. Treasury ETFs are still the biggest type in our sample: With about US$50 billion AuM, they constitute about half of the total of US$91 billion. Next are investment-grade ETFs, with AuM of about US$32 billion. High-yield ETFs come third, with US$9 billion. Note that our sample covers 4% of the total AuM of the fixed-income ETF market. In the last column of Exhibit 1, we present the average TERs. Treasury ETFs' average TER is the lowest at 16 bps, followed closely by investment-grade ETFs with an average TER of 18 bps. High-yield ETFs are about three times as expensive: 47 bps. These average TERs are relevant because they provide a yardstick for the underperformance of ETFs, as argued earlier.

EXHIBIT 1
Number of ETFs, Their AuM (in US$ billions), and Their Average TER (in bps) by Benchmark Type

	Number	AuM	TER
Treasury	106	49.6	16
Investment Grade	20	31.9	18
High Yield	3	9.2	47
Total	129	90.7	-

Source: BlackRock [2010a].

RESULTS

In this section, we test our hypotheses using empirical data. For every ETF and for every date, we calculate the difference between its net total return and its benchmark's total return. We calculate the average of these differences over the period from the second full month after an ETF's inception[6] until July 30, 2010 (the last date in our sample). We calculate an ETF's outperformance as the annualized, average total return difference. Further, we define an ETF's cost-adjusted outperformance as its outperformance plus its TER.

Exhibit 2 presents the average outperformance by benchmark type. For each benchmark type, we calculate the equally weighted average outperformance and

EXHIBIT 2
Average Outperformance and Cost-Adjusted Outperformance (in bps) by Benchmark Type

	Outperformance	Cost-Adjusted Outperformance
Treasury	–16	0
Investment Grade	–56	–38
High Yield	–384	–337

cost-adjusted outperformance for all ETFs with that benchmark type.[7] These calculations allow us to test hypotheses H1, H2, and H3.

For Treasury ETFs, the average cost-adjusted outperformance equals 0. Hence, Treasury ETFs are indeed able to track their benchmarks before costs and lag their benchmark by their TER after costs. This result is consistent with hypothesis H1.

For investment-grade and high-yield corporate bond ETFs, the average cost-adjusted outperformance is negative, as hypothesized in H2. So, because of corporate bonds' higher trading costs relative to Treasury bonds, corporate bond ETFs lag their benchmarks by more than their costs.

In addition, the cost-adjusted outperformance is more negative for high yield than for investment grade. This reflects the higher transaction costs for high-yield bonds compared to investment-grade bonds. This validates hypothesis H3. The costs of tracking a corporate bond benchmark, over and above the TER, amount to 38 bps for investment grade and 337 bps for high yield.

To test hypotheses H4 and H5, we create several subgroups of investment-grade ETFs, as shown in Exhibit 3.[8] Panel A shows that ETFs benchmarked to an index consisting of bonds longer than five years have larger underperformances than those benchmarked to an index consisting of bonds shorter than five years. This is consistent with hypothesis H4 that says that longer-maturity corporate bonds have larger trading costs than shorter-dated bonds.

EXHIBIT 3

Average Outperformance and Cost-Adjusted Outperformance (in bps) by Maturity and Benchmark Liquidity Type of Investment-Grade ETFs

	Outperformance	Cost-Adjusted Outperformance
Panel A: Benchmark Maturity		
Less Than Five Years	–43	–25
Five Years or More	–106	–89
Panel B: "Liquid" Benchmark		
Yes	–28	–10
No	–67	–50

Panel B shows that the average underperformance of ETFs with a "liquid" benchmark is indeed smaller than of ETFs with "non-liquid" benchmarks, as expected in hypothesis H5. A benchmark consisting of a smaller number of larger bonds is easier to track than more broadly defined benchmarks that also contain smaller issues.

CONCLUSIONS AND IMPLICATIONS

In this article, we analyze the performance of ETFs on Treasury and corporate bond indices. Our main findings are as follows:

- Because Treasury bonds are more liquid and cheaper to trade than corporate bonds, Treasury ETFs can track their benchmarks but corporate bond ETFs underperform their benchmarks;
- Because investment-grade corporate bonds have lower transaction costs than high-yield corporate bonds, the underperformance of high-yield ETFs exceeds that of investment-grade ETFs;
- Because shorter-maturity bonds have smaller bid–ask spreads than longer-maturity bonds, investment-grade ETFs with shorter-maturity benchmarks underperform less than those with longer-maturity benchmarks; and
- Because larger bonds have smaller bid–ask spreads than smaller bonds, investment-grade ETFs that are benchmarked to "liquid" indices, consisting of a smaller number of larger bonds, underperform less than ETFs benchmarked to a broad index.

These results indicate that an ETF's ability to replicate its benchmark depends on the transaction costs of that benchmark's constituents. The more costly it is to trade the constituents, the harder it is to replicate the benchmark's returns.

Our results are consistent with the empirical literature on transaction costs and the liquidity of government bonds and corporate bonds. Levine, Drucker, and Rosenthal [2010] also recently observed that index tracking is problematic for high-yield benchmarks. We find that these problems also carry over, although to a lesser extent, to investment-grade corporate bonds indices.

The inability of passive corporate bond ETFs to replicate their benchmarks' returns also has implications for measuring the added value of active managers. For instance, Lipton and Kish [2010] concluded that high-

yield fund managers added little return in excess of the Barclays Capital U.S. High Yield Index. Our results, however, show that passively managed high-yield ETFs have also been unable to replicate the returns of their benchmarks. Therefore, the benchmark return does not seem to be a fair yardstick for evaluating active managers' added value. Instead, the return of passively managed funds, such as ETFs, could be used to benchmark the returns of actively managed funds.

A second implication of our study is for investors making asset allocation decisions and implementing those decisions with ETFs. Given that corporate bond ETFs lag their benchmarks by more than their costs, investors in ETFs should lower their return expectations for investment-grade and high-yield corporate bonds. Investors cannot simply use the historical mean return of the benchmarks, because the ETFs are unable to generate those benchmark returns. Instead, investors should apply a discount that depends on the benchmark constituents' transaction costs.

ENDNOTES

I thank David Blitz and Martin Martens for comments on an earlier version of this article. Any remaining errors are my own. Views expressed in this article are my own and do not necessarily reflect those of Robeco.

[1]This section draws on BlackRock [2010a] and wikipedia.org/wiki/Exchange_traded_funds

[2]Actively managed ETFs also exist. Further, leveraged ETFs and inverse ETFs (or a combination of both) have been introduced.

[3]Other commonly used expressions are relative performance, tracking error, or tracking difference.

[4]Other determinants, such as trading volume, bond age, and coupon, cannot be tested, either because the data are unavailable or because we analyze returns on the *fund* level data instead of the *bond* level.

[5]The difference between credits and corporate bonds is that the latter includes only corporations, whereas the former also includes noncorporate issuers such as government-related entities.

[6]For each ETF, the first month after inception is skipped, because the returns in this first month were often erratic.

[7]We bear in mind that these outperformances are based on data histories with different lengths.

[8]This is unfortunately not possible for high-yield ETFs, because there are only three of them. For Treasury ETFs, we found no remarkable patterns after analyzing subgroups, and these results thus are not reported.

REFERENCES

BlackRock. ETF Landscape Q2 2010, 2010a.

——. ETF Landscape Q3 2010, 2010b.

Blitz, D.C., and J. Huij. "Evaluating the Performance of Global Emerging Markets Equity Exchange-Traded Funds." Working paper, Erasmus University, 2011.

Blitz, D.C., J. Huij, and L.A.P. Swinkels. "The Performance of European Index Funds and ETFs." *European Financial Management*, forthcoming 2012.

Chakravarty, S., and A. Sarkar. "Trading Costs in Three U.S. Bond Markets." *The Journal of Fixed Income*, June 2003, pp. 39-48.

Chen, L., D.A. Lesmond, and J. Wei. "Corporate Yield Spreads and Bond Liquidity." *Journal of Finance*, Vol. 62, No. 1 (2007), pp. 229-249.

Derwall, J., and J. Huij. "'Hot Hands' in Bond Funds." *Journal of Banking and Finance*, Vol. 32, No. 4 (2008), pp. 559-572.

Drenovak, M., B. Uroševic, and R. Jelic. "European Bond ETFs—Tracking Errors and Sovereign Debt Crisis." Working paper, University of Birmingham, 2010.

Houweling, P., A.A. Mentink, and A.C.F. Vorst. "Comparing Possible Proxies of Corporate Bond Liquidity." *Journal of Banking and Finance*, Vol. 29, No. 6 (2005), pp. 1331-1358.

Levine, R., E. Drucker, and S. Rosenthal. "The Problems and Challenges of High-Yield Bond Benchmarking." *The Journal of Portfolio Management* (Summer 2010), pp. 93-98.

Lipton, A.F., and R.J. Kish. "Robust Performance Measures for High Yield Bond Funds." *Quarterly Review of Economics and Finance*, Vol. 50, No. 3 (2010), pp. 332-340.

Rompotis, G.G. "Performance and Trading Characteristics of German Passively Managed ETFs." *International Research Journal of Finance and Economics*, Vol. 15 (2008), pp. 210-223.

Tucker, M., and S. Laipply. "Understanding Bond ETF Premiums and Discounts: A Conceptual Framework." *Journal of Indexes* (September/October 2010), pp. 40-48.